Document made available under the Patent Cooperation Treaty (PCT)

International application number: PCT/EP04/014170

International filing date: 13 December 2004 (13.12.2004)

Document type: Certified copy of priority document

Document details: Country/Office: EP
Number: 04027213.0
Filing date: 16 November 2004 (16.11.2004)

Date of receipt at the International Bureau: 28 January 2005 (28.01.2005)

Remark: Priority document submitted or transmitted to the International Bureau in compliance with Rule 17.1(a) or (b)

Europäisches Patentamt
European Patent Office
Office européen des brevets

Bescheinigung Certificate Attestation

Die angehefteten Unterlagen stimmen mit der ursprünglich eingereichten Fassung der auf dem nächsten Blatt bezeichneten europäischen Patentanmeldung überein.

The attached documents are exact copies of the European patent application described on the following page, as originally filed.

Les documents fixés à cette attestation sont conformes à la version initialement déposée de la demande de brevet européen spécifiée à la page suivante.

Patentanmeldung Nr. Patent application No. Demande de brevet n°

04027213.0

Der Präsident des Europäischen Patentamts;
Im Auftrag

For the President of the European Patent Office

Le Président de l'Office européen des brevets
p.o.

R C van Dijk

Europäisches Patentamt | European Patent Office | Office européen des brevets

Anmeldung Nr:
Application no.: 04027213.0
Demande no:

Anmeldetag:
Date of filing: 16.11.04
Date de dépôt:

Anmelder/Applicant(s)/Demandeur(s):

Epigenomics AG
Kastanienallee 24
10435 Berlin
ALLEMAGNE

Bezeichnung der Erfindung/Title of the invention/Titre de l'invention:
(Falls die Bezeichnung der Erfindung nicht angegeben ist, siehe Beschreibung.
If no title is shown please refer to the description.
Si aucun titre n'est indiqué se referer à la description.)

PIXT2 - a marker to predict survival of patients diagnosed with breast cell proliferative disease

In Anspruch genommene Priorität(en) / Priority(ies) claimed /Priorité(s) revendiquée(s)
Staat/Tag/Aktenzeichen/State/Date/File no./Pays/Date/Numéro de dépôt:

Internationale Patentklassifikation/International Patent Classification/
Classification internationale des brevets:

G01N33/48

Am Anmeldetag benannte Vertragstaaten/Contracting states designated at date of filing/Etats contractants désignées lors du dépôt:

AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HU IE IS IT LU MC
NL PL PT RO SE SI SK TR LI

FB14632

PITX2 - a marker to predict survival of patients diagnosed with breast cell proliferative disease

The present invention relates to methods for predicting the survival of a human being diagnosed with a cell proliferative disorder of the breast tissues, characterised by a step of determining the expression level of PITX2 or the genetic or the epigenetic modifications of the genomic DNA associated with the gene PITX2. The invention also relates to sequences, oligonucleotides and antibodies which can be used within the described methods.

Field of the Invention

BREAST CANCER SURVIVAL

In European and American women breast cancer is the most frequently diagnosed cancer and the second leading cause of cancer death. In women aged 40-5, breast cancer is the leading cause of death (Greenlee *et al.*, 2000). In 2002 there were 204,000 new cases of breast cancer in the US and a comparable number in Europe.

Breast cancer is defined as the uncontrolled proliferation of cells within breast tissues. Breasts are comprised of 15 to 20 lobes joined together by ducts. Cancer arises most commonly in the duct, but is also found in the lobes with the rarest type of cancer termed inflammatory breast cancer. It will be appreciated by those skilled in the art that there exists a continuing need to improve methods of early detection, classification and treatment of breast cancers. In contrast to the detection of some other common cancers such as cervical and dermal there are inherent difficulties in classifying and detecting breast cancers.

The first step of any treatment is the assessment of the patient's condition comparative to defined classifications of the disease. However the value of such a system is inherently dependent upon the quality of the classification. Breast cancers are staged according to their size, location and occurrence of metastasis. Methods of treatment include the use of surgery, radiation therapy, chemotherapy and endocrine therapy, which are also used as adjuvant therapies to surgery. In general more aggressive diseases should be treated with more aggressive therapies.

Although the vast majority of early cancers are operable, i.e. the tumor can be completely removed by surgery, about one third of the patients with lymph-node negative diseases and about 50-60% of patients with node-positive disease will develop metastases during follow-up.

Based on this observation, systemic adjuvant treatment has been introduced for both node-positive and node-negative breast cancers. Systemic adjuvant therapy is administered after surgical removal of the tumour, and has been shown to reduce the risk of recurrence significantly. Several types of adjuvant treatment are available: endocrine treatment, also called hormone treatment (for hormone receptor positive tumours), different chemotherapy regimens, and antibody treatments based on novel agents like Herceptin (an antibody to an epidermal growth factor receptor).

The growth of the majority of breast cancers (app. 70-80%) is dependent on the presence of estrogen. Therefore, one important target for adjuvant therapy is the removal of estrogen (e.g. by ovarian ablation) or the blocking of its synthesis or the blocking of its actions on the tumour cells either by blocking the receptor with competing substances (e.g. Tamoxifen) or by inhibiting the conversion of androgen into estrogen (e.g. aromatase inhibitors). This type of treatment is called "endocrine treatment". Endocrine treatment is thought to be efficient only in tumours that express hormone receptors (the estrogen receptor (ER) and/or the progesterone receptor (PR)). Currently, the vast majority of women with hormone receptor positive breast cancer receive some form of endocrine treatment, independent of their nodal status. The most frequently used drug in this scenario is Tamoxifen.

However, even in hormone receptor positive patients, not all patients benefit from endocrine treatment. Adjuvant endocrine therapy reduces mortality rates by 22% while response rates to endocrine treatment in the metastatic (advanced) setting are 50 to 60%.

Since Tamoxifen has relatively few side effects, treatment may be justified even for patients with low likelihood of benefit. However, these patients may require additional, more aggressive adjuvant treatment. Even in earliest and least aggressive tumours, such as node-negative, hormone receptor positive tumours, about 21% of patients relapse within 10 years after initial diagnosis if they receive Tamoxifen monotherapy only, as adjuvant treatment (Lancet. 1998

May 16;351(9114):1451-67. Tamoxifen for early breast cancer: an overview of the randomised trials. Early Breast Cancer Trialists' Collaborative Group.). Similarly, some patients with hormone receptor negative disease may be treated sufficiently with surgery and potentially radiotherapy alone, whereas others may require additional chemotherapy.

Several cytotoxic regimens have shown to be effective in reducing the risk of relapse in breast cancer (Mansour *et al.*, Survival advantage of adjuvant chemotherapy in high-risk node-negative breast cancer: ten-year analysis--an intergroup study. J Clin Oncol. 1998 Nov;16(11):3486-92.). According to current treatment guidelines, most node-positive patients receive adjuvant chemotherapy both in the US and Europe, since the risk of relapse is considerable. Nevertheless, not all patients do relapse, and there is a proportion of patients who would never have relapsed even without chemotherapy, but who nevertheless receive chemotherapy due to the currently used criteria. In hormone receptor positive patients, chemotherapy is usually given before endocrine treatment, whereas hormone receptor negative patients receive only chemotherapy.

The situation for node-negative patients is particularly complex. In the US, cytotoxic chemotherapy is recommended for node-negative patients, if the tumour is larger than 1 cm. In Europe, chemotherapy is considered for the node-negative cases if one or more risk factors such as tumour size larger than 2 cm, negative hormone receptor status, or tumour grading of three or age <35 is present. In general, there is a tendency to select premenopausal women for additional chemotherapy whereas for postmenopausal women, chemotherapy is often omitted. Compared to endocrine treatment, in particular Tamoxifen or aromatase inhibitors, chemotherapy is highly toxic, with short-term side effects such as nausea, vomiting, bone marrow depression, and long-term effects such as cardiotoxicity and an increased risk for secondary cancers.

It is currently not clear which breast cancer patients should be selected for more aggressive therapy and which would do well without additional aggressive treatment, and clinicians agree that there is a large need for proper selection of patients. The difficulty of selecting the right patients for adjuvant treatment and selecting the right adjuvant treatment, and the lack of suitable criteria is also reflected by a recent study which showed that chemotherapy is used much less frequently than recommended, based on data from the New Mexico Tumor registry

(Du *et al.*, 2003). This study provided substantial evidence that there is a need for better selection of patients for chemotherapy or other, more aggressive forms of breast cancer therapy.

PITX2 (also known as PTX2, RS, RGS, ARP1, Brx1, IDG2, IGDS, IHG2, RIEG, IGDS2, IRID2, Otlx2, RIEG1, MGC20144) is known to belong to the PTX subfamily of PTX1, PTX2, and PTX3 genes which define a novel family of transcription factors, within the paired-like class of homeodomain factors. The gene PITX2 (according to NM_153426) encodes the paired-like homeodomain transcription factor 2, which is known to be expressed during development of anterior structures such as the eye, teeth, and anterior pituitary.

Toyota et *al.*, (2001) (Blood 97: p 2823-9.) found hypermethylation of the PITX2 gene in a large proportion of acute myeloid leukemias. Furthermore, in this study hypermethylation of PITX2 is positively correlated to methylation of the ER gene and to a reduced expression level. Means to analyse the methylation pattern of the PITX2 gene have been described in a number of patent applications, too (WO 02/077272 is related to the use of methylation markers to differentiate between AML and ALL, WO 01/19845 is related to several differentially methylated sequences useful for diagnosis of several cell proliferative disorders, WO 02/00927 and WO 01/092565 are related to the use of methylation markers to diagnose diseases associated with development genes or associated with DNA transcription, respectively.
Loss of heterozygosity (hereinafter also referred to as 'LOH') of chromosome 4 is a known characteristic of many tumour types. Shivapurkar et al. [*Cancer Research* 59, 3576-3580, August 1, 1999] have observed loss of heterozygosity at multiple regions of chromosome 4 in breast cancer samples and cell lines. Deletions at 4q25–26 were present in 67% of analysed samples. However the analysed region (between markers D4S1586 and D4S175) does not map to the PITX2 gene, and no inference concerning PITX2 expression was made. Furthermore, the investigation as carried out does not indicate the suitability of any genes or loci of the region for a prognostic use.

Although the methylation of PITX2 has been associated with development, transcription and disease such as cancer, it has no heretofore recognised role in the outcome prediction of breast cancer patients or responsiveness to endocrine treatment.

EXPRESSION ANALYSIS

The expression of a gene, or rather the protein encoded by the gene, can be studied on four different levels: firstly, protein expression levels can be determined directly, secondly, mRNA transcription levels can be determined, thirdly, epigenetic modifications, such as gene's DNA methylation profile or the gene's histone profile; can be analysed, as methylation is often correlated with inhibited protein expression, and fourth, the gene itself may be analysed for genetic modifications such as mutations, deletions, polymorphisms etc. influencing the expression of the gene product.

The levels of observation that have been studied by the methodological developments of recent years in molecular biology, are the genes themselves, the transcription of these genes into RNA, and the translation into the resulting proteins. However how the activation and inhibition of specific genes, in specific cells and tissues, at specific time points in the course of development of an individual are controlled, is correlatable to the degree and character of the methylation of the genes or respectively the genome. In this respect, pathogenic conditions may manifest themselves in a changed methylation pattern of individual genes or of the genome.

The four terms that apply to the fields of overall genome-wide analysis of all these biological processes are called: Proteomics, Transcriptomics, Epigenomics (or Methylomics) and Genomics. Methods and techniques that can be used for studying expression or studying the modifications responsible for expression on all of these levels are well described in the literature and therefore known to a person skilled in the art. They are described in text books of molecular biology and in a large number of scientific journals.

How to analyse the protein expression of a single gene is prior art. It usually requires an antibody specific for the gene product of interest. Appropriate technologies would be ELISA or Immunohistochemistry.

The analysis of the level of mRNA also has been described sufficiently. These days the gold standard is the reverse transcriptase PCR.

To avoid duplication a more detailed description of the prior art relating to existing and well known technologies is given within the description of the invention, as it is part of the invention.

US patent application 2003/0198970 by Gareth Roberts lists some of the technologies and methods on how to determine a person's "genetic make up", i.e. the genetic modifications, such as deletions, polymorphisms, mutations etc. that may vary between individuals and describes the potential role of this genetic sequence information in the individual's variability in disease, response to therapy and prognosis. Epigenetic differences however are not mentioned. The gene PITX2 is listed within this application as one gene name out of a long and comprehensive list of about 2.500 other gene names, suggesting its expression could play a role in some kind of treatment response. However, this is simply an assumption based on speculation only, as no experiments are disclosed, which demonstrate any kind of relation between genetic modifications of PITX2 and an individual's variation in treatment response.

A less established area in this context is the field of epigenomics or epigenetics, i.e. the field concerned with analysis of DNA methylation patterns.

Methylation of DNA can play an important role in the control of gene expression in mammalian cells. DNA methyltransferases are involved in DNA methylation and catalyse the transfer of a methyl group from S-adenosylmethionine to cytosine residues to form 5-methylcytosine, a modified base that is found mostly at CpG sites in the genome. The presence of methylated CpG islands in the promoter region of genes can suppress their expression. This process may be due to the presence of 5-methylcytosine, which apparently interferes with the binding of transcription factors or other DNA-binding proteins to block transcription. In different types of tumours, aberrant or accidental methylation of CpG islands in the promoter region has been observed for many cancer-related genes, resulting in the silencing of their expression. Such genes include tumour suppresser genes, genes that suppress metastasis and angiogenesis, and genes that repair DNA (Momparler and Bovenzi (2000) J. Cell Physiol. 183:145-54).

In addition it has been described that DNA methylation may also play a role in the field of pharmacogenetics. A similar approach on how to apply information about genetic modifications of the genome to the analysis of individual responses to treatment as was for example described by Gareth Roberts in US application 2003/0198970 was already subject of the application WO 02/037398, tailored to the application of information about *epigenetic* modifications of the genome, based on DNA methylation analysis, to guide treatment selection and to study individual's treatment responses.

An example for the applicability of this idea was given by Esteller et al. (Esteller et al. (2000) N Engl J Med. 2000 Nov 9;343(19):1350-4.), who demonstrated that methylation of the MGMT promoter in gliomas is a useful predictor of the responsiveness of the tumours to alkylating agents. More recently, Frühwald has summarised a series of studies demonstrating that DNA methylation is associated with the aggressiveness of different cancers (Fruhwald MC. DNA methylation patterns in cancer: novel prognostic indicators? Am J Pharmacogenomics. 2003;3(4):245-60).

An example for the potential of analysis of epigenetic modifications, such as DNA methylation analysis, for the prediction of treatment response - related to breast cancer- was presented as a poster by Martens et al. at the San Antonio Breast Cancer Symposium, San Antonio, TX, December, 3-6, 2003. Breast cancer patients which have had their tumours removed by surgery and developed metastases at some point after the removal, were treated with Tamoxifen, an endocrine treatment drug. The primary tumour samples were analysed for aberrant methylation patterns. The patients were then divided into two sub classes according to their objective tumour response: patients with progressive disease (which could be described as increasing metastasis size) and patients with complete or partial remission of the relapsed tumour (which could be described as decreasing metastasis size). It turned out, that those patients which had a tumour removed and experienced a remission (decrease in size) of the metastasis under endocrine treatment, had suffered from a tumour which showed a distinct pattern of DNA methylation at specific CpG sites, whereas patients which show progressive disease (did not experience a decrease but an increase in size of their metastases), under endocrine treatment, suffered from a tumour which did not show this distinct pattern of DNA methylation (but a different pattern) at these CpG sites. This is a clear indication, that the methylation pattern described in that study can serve as a **predictive** treatment response tool for an endocrine treatment, like tamoxifen. The results of this study, i.e. predictive biomarkers and assays

therefore, are subject of patent application WO 04/035803, published at April 29, 2004: Method and nucleic acid for the improved treatment of breast cell proliferative disorders. Predictive markers as described above will also be called 'metastatic' markers in the context of this application. PITX2 is also listed as a predictive marker in said application.

Currently several predictive markers are under evaluation. As up to now most patients have received Tamoxifen as endocrine treatment most of the markers have been shown to be associated with response or resistance to tamoxifen. However, it is generally assumed that there is a large overlap between responders to one or the other endocrine treatment. In fact, ER and PR expression are used to select patients for any endocrine treatment. Among the markers which have been associated with tamoxifen response is bcl-2. High bcl-2 expression levels showed promising correlation to tamoxifen therapy response in patients with metastatic disease and prolonged survival and added valuable information to an ER negative patient subgroup (J Clin Oncology, 1997, 15 5: 1916-1922; Endocrine, 2000, 13(1):1-10). There is conflicting evidence regarding the independent predictive value of c-erbB2 (Her2/neu) overexpression in patients with advanced breast cancer that require further evaluation and verification (British J of Cancer, 1999, 79 (7/8):1220-1226; J Natl Cancer Inst, 1998, 90 (21): 1601-1608).

Other predictive markers include SRC-1 (steroid receptor coactivator-1), CGA mRNA over expression, cell kinetics and S phase fraction assays (Breast Cancer Res and Treat, 1998, 48:87-92; Oncogene, 2001, 20:6955-6959). Recently, uPA (Urokinase-type plasminogen activator) and PAI-1 (Plasminogen activator inhibitor type 1) together showed to be useful to define a subgroup of patients who have worse prognosis and who would benefit from adjuvant systemic therapy (J Clinical Oncology, 2002, 20 n° 4). However, all of these markers need further evaluations in prospective trials as none of them is yet a validated marker of response.

Also recently published was a study related to the **prognostic** power of methylation analysis in breast cancer patients. Müller et al. (Muller HM, Widschwendter A, Fiegl H, Ivarsson L, Goebel G, Perkmann E, Marth C, Widschwendter M. (2003) DNA methylation in serum of breast cancer patients: an independent prognostic marker. Cancer Res. 2003 Nov 15; 63(22): 7641-5.) reported about a set of genes, which can be used as biomarkers in patient pretherapeutic sera for the prognosis of breast cancer. Specific aberrant methylation patterns of

two genes found in DNA from pretreatment serum of cancer patients indicated whether their prognosis was good or bad. The DNA analysed was not tumour DNA but serum DNA. Most likely the presence of a tumour-specific pattern indicates that tumour derived DNA is present, however, the absence of a specific methylation pattern, may be due to a tumour which does not show this methylation pattern, or a tumour which does not shed sufficient DNA into the blood stream. Good or bad prognosis was defined as long or short "overall survival" after surgery, without adjuvant treatment. This result therefore relates to untreated patients, only.

These 'prognostic' markers are able to answer the question whether or not a breast cancer patient should get an aggressive adjuvant treatment like chemotherapy after removal of the tumour to avoid recurrence of cancer, i.e. occurrence of metastases.

However, none of these study results and none of these markers is able to answer the specific question raised above, whether or not a breast cancer patient should get adjuvant chemotherapy after removal of the tumour to avoid recurrence of cancer, i.e. occurrence of metastases in addition or alternative to endocrine treatment (with a drug like tamoxifen, or aromatase inhibitors).

A marker for a bad prognosis for cancer patients (without treatment), might not be applicable to a patient under adjuvant treatment with a drug like tamoxifen. Therefore the test would not be able to help deciding, whether chemotherapy, including all its side affects and inherent risks, is necessary or whether endocrine treatment is sufficient, because an endocrine treatment might change the prognosis from "bad" to "good".

The predictive 'metastatic' marker set described above, would be able to identify amongst all patients which relapsed (developed metastases after surgery) those patients, which do not respond to endocrine treatment (by partial or complete remission of relapsed tumour). These markers however, cannot be applied to answer the question whether metastases will occur at all (after surgery of the primary tumour under endocrine treatment), and consequently whether it is advised to give adjuvant chemotherapy to avoid recurrence of cancer (i.e. relapse or occurrence of metastases).

In one aspect the present invention provides a marker, PITX2 (which shall be recognised as the gene encoding for the protein PITX2; according to NM_153426), that can be used to an-

swer that question and help guiding the decision whether or not an adjuvant chemotoxic therapy shall be subscribed in addition or instead of treatment with endocrines, like tamoxifen. A marker able to answer this question will also be called 'adjuvant' marker, in the context of this application.

The herein described invention provides a novel breast cell proliferative disorder prognostic biomarker.

It is herein disclosed that aberrant expression of the gene PITX2 is correlated to prognosis of breast cell proliferative disorder patients, in particular breast carcinoma.

In particular this marker provides a novel means for the characterisation of breast carcinomas. Aberrant expression of the gene PITX2 is indicative of the survival of a breast carcinoma patient treated with one or more treatments which target the estrogen receptor, synthesis or conversion pathways or are otherwise involved in estrogen metabolism, production or secretion. The herein described invention is particularly useful for the differentiation of individuals who may be appropriately treated with one or more treatments which target the estrogen receptor pathway or are involved in estrogen metabolism, production or secretion from those individuals who would be optimally treated with other treatments in addition to said treatment. Preferred 'other treatments' include but are not limited to chemotherapy or radiotherapy.

As used herein the term expression shall be taken to mean the transcription and translation of a gene. The level of expression of a gene may be determined by the analysis of any factors associated with or indicative of the level of transcription and translation of a gene including but not limited to methylation analysis, loss of heterozygosity (hereinafter also referred to as LOH), RNA expression levels and protein expression levels.

Furthermore the activity of the transcribed gene may be affected by genetic variations such as but not limited genetic mutations (including but not limited to SNPs, point mutations, deletions, insertions, repeat length, rearrangements and other polymorphisms).

In addition study results presented by Paik et al. at the San Antonio Breast Cancer Symposium, San Antonio, TX, December, 3-6, 2003 provide an answer to this question, by analysing the mRNA expression pattern of 16 genes plus 5 controls with RT-PCR.

The provided invention however has the advantage that looking at only one gene or a small selection of three to five genes will give sufficient information for a validated prognosis.

For demonstration : The 'metastatic' test (use of a 'metastatic' marker) tells a patient whether she is unlikely to respond to endocrine treatment when she develops metastases. But she does not know how high the likelihood is, that she will experience a relapse at all. The 'prognostic' test (use of a 'prognostic' marker) tells a patient whether she will have a good or bad prognosis without any treatment. Even with a "bad prognosis" endocrine treatment might be enough though. The prognostic markers are not necessarily able to predict the outcome under endocrine treatment. The 'adjuvant test' (use of an 'adjuvant' marker) tells her whether she will or will not develop recurrence, without chemotherapy, even when treated with the standard -low side effected- endocrine treatment.

This invention relates to the use of PITX2, as an 'adjuvant marker', which also serves as a 'prognostic marker', especially in hormone receptor negative women, which would not get any endocrine treatment at all.

5-methylcytosine is the most frequent covalent base modification in the DNA of eukaryotic cells. It plays a role, for example, in the regulation of the transcription, in genetic imprinting, and in tumorigenesis. Therefore, the identification of 5-methylcytosine as a component of genetic information is of considerable interest. However, 5-methylcytosine positions cannot be identified by sequencing since 5-methylcytosine has the same base pairing behaviour as cytosine. Moreover, the epigenetic information carried by 5-methylcytosine is completely lost during PCR amplification.

A relatively new and currently the most frequently used method for analysing DNA for 5-methylcytosine is based upon the specific reaction of bisulfite with cytosine which, upon subsequent alkaline hydrolysis, is converted to uracil which corresponds to thymidine in its base pairing behaviour. However, 5-methylcytosine remains unmodified under these conditions. Consequently, the original DNA is converted in such a manner that methylcytosine, which originally could not be distinguished from cytosine by its hybridisation behaviour, can now be detected as the only remaining cytosine using "normal" molecular biological techniques, for example, by amplification and hybridisation or sequencing. All of these techniques are based

on base pairing which can now be fully exploited. In terms of sensitivity, the prior art is defined by a method which encloses the DNA to be analysed in an agarose matrix, thus preventing the diffusion and renaturation of the DNA (bisulfite only reacts with single-stranded DNA), and which replaces all precipitation and purification steps with fast dialysis (Olek A, Oswald J, Walter J. A modified and improved method for bisulphite based cytosine methylation analysis. Nucleic Acids Res. 1996 Dec 15;24(24):5064-6). Using this method, it is possible to analyse individual cells, which illustrates the potential of the method. However, currently only individual regions of a length of up to approximately 3000 base pairs are analysed, a global analysis of cells for thousands of possible methylation events is not possible. However, this method cannot reliably analyse very small fragments from small sample quantities either. These are lost through the matrix in spite of the diffusion protection.

An overview of the further known methods of detecting 5-methylcytosine may be gathered from the following review article: Rein, T., DePamphilis, M. L., Zorbas, H., Nucleic Acids Res. 1998, 26, 2255.

To date, barring few exceptions (e.g., Zeschnigk M, Lich C, Buiting K, Doerfler W, Horsthemke B. A single-tube PCR test for the diagnosis of Angelman and Prader-Willi syndrome based on allelic methylation differences at the SNRPN locus. Eur J Hum Genet. 1997 Mar-Apr;5(2):94-8) the bisulfite technique is only used in research. Always, however, short, specific fragments of a known gene are amplified subsequent to a bisulfite treatment and either completely sequenced (Olek A, Walter J. The pre-implantation ontogeny of the H19 methylation imprint. Nat Genet. 1997 Nov;17(3):275-6) or individual cytosine positions are detected by a primer extension reaction (Gonzalgo ML, Jones PA. Rapid quantitation of methylation differences at specific sites using methylation-sensitive single nucleotide primer extension (Ms-SNuPE). Nucleic Acids Res. 1997 Jun 15;25(12):2529-31, WO 95/00669) or by enzymatic digestion (Xiong Z, Laird PW. COBRA: a sensitive and quantitative DNA methylation assay. Nucleic Acids Res. 1997 Jun 15;25(12):2532-4). In addition, detection by hybridisation has also been described (Olek et al., WO 99/28498).

Further publications dealing with the use of the bisulfite technique for methylation detection in individual genes are: Grigg G, Clark S. Sequencing 5-methylcytosine residues in genomic DNA. Bioessays. 1994 Jun;16(6):431-6, 431; Zeschnigk M, Schmitz B, Dittrich B, Buiting K, Horsthemke B, Doerfler W. Imprinted segments in the human genome: different DNA meth-

ylation patterns in the Prader-Willi/Angelman syndrome region as determined by the genomic sequencing method. Hum Mol Genet. 1997 Mar;6(3):387-95; Feil R, Charlton J, Bird AP, Walter J, Reik W. Methylation analysis on individual chromosomes: improved protocol for bisulphite genomic sequencing. Nucleic Acids Res. 1994 Feb 25;22(4):695-6; Martin V, Ribieras S, Song-Wang X, Rio MC, Dante R. Genomic sequencing indicates a correlation between DNA hypomethylation in the 5' region of the pS2 gene and its expression in human breast cancer cell lines. Gene. 1995 May 19;157(1-2):261-4; WO 97/46705 and WO 95/15373.

An overview of the Prior Art in oligomer array manufacturing can be gathered from a special edition of Nature Genetics (Nature Genetics Supplement, Volume 21, January 1999), published in January 1999, and from the literature cited therein.

Fluorescently labelled probes are often used for the scanning of immobilised DNA arrays. The simple attachment of Cy3 and Cy5 dyes to the 5'-OH of the specific probe are particularly suitable for fluorescence labels. The detection of the fluorescence of the hybridised probes may be carried out, for example via a confocal microscope. Cy3 and Cy5 dyes, besides many others, are commercially available.

Matrix Assisted Laser Desorption Ionization Mass Spectrometry (MALDI-TOF) is a very efficient development for the analysis of biomolecules (Karas M, Hillenkamp F. Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. Anal Chem. 1988 Oct 15;60(20):2299-301). An analyte is embedded in a light-absorbing matrix. The matrix is evaporated by a short laser pulse thus transporting the analyte molecule into the vapour phase in an unfragmented manner. The analyte is ionised by collisions with matrix molecules. An applied voltage accelerates the ions into a field-free flight tube. Due to their different masses, the ions are accelerated at different rates. Smaller ions reach the detector sooner than bigger ones.

MALDI-TOF spectrometry is excellently suited to the analysis of peptides and proteins. The analysis of nucleic acids is somewhat more difficult (Gut I G, Beck S. DNA and Matrix Assisted Laser Desorption Ionization Mass Spectrometry. Current Innovations and Future Trends. 1995, 1; 147-57). The sensitivity to nucleic acids is approximately 100 times worse than to peptides and decreases disproportionally with increasing fragment size. For nucleic

acids having a multiply negatively charged backbone, the ionisation process via the matrix is considerably less efficient. In MALDI-TOF spectrometry, the selection of the matrix plays an eminently important role. For the desorption of peptides, several very efficient matrixes have been found which produce a very fine crystallisation. There are now several responsive matrixes for DNA, however, the difference in sensitivity has not been reduced. The difference in sensitivity can be reduced by chemically modifying the DNA in such a manner that it becomes more similar to a peptide. Phosphorothioate nucleic acids in which the usual phosphates of the backbone are substituted with thiophosphates can be converted into a charge-neutral DNA using simple alkylation chemistry (Gut IG, Beck S. A procedure for selective DNA alkylation and detection by mass spectrometry. Nucleic Acids Res. 1995 Apr 25;23(8):1367-73). The coupling of a charge tag to this modified DNA results in an increase in sensitivity to the same level as that found for peptides. A further advantage of charge tagging is the increased stability of the analysis against impurities which make the detection of unmodified substrates considerably more difficult.

Genomic DNA is obtained from DNA of cell, tissue or other test samples using standard methods. This standard methodology is found in references such as Sambrook, Fritsch and Maniatis eds., Molecular Cloning: A Laboratory Manual, 1989.

DESCRIPTION

Characterisation of a breast cancer in terms of its predicted aggressiveness enables the physician to make an informed decision as to a therapeutic regimen with appropriate risk and benefit trade offs to the patient.

Aggressiveness is taken to mean one or more of decreased patient survival or disease- or relapse-free survival, increased tumour-related complications and faster progression of tumour or metastases. According to the aggressiveness of the disease an appropriate treatment or treatments may be selected from the group consisting of chemotherapy, radiotherapy, surgery, biological therapy, immunotherapy, antibody treatments, treatments involving molecularly targeted drugs, estrogen receptor modulator treatments, estrogen receptor down-regulator treatments, aromatase inhibitors treatments, ovarian ablation, treatments providing LHRH analogues or other centrally acting drugs influencing estrogen production. Wherein a cancer is

characterised as 'aggressive' it is particularly preferred that a treatment such as, but not limited to, chemotherapy is provided in addition to or instead of an endocrine targeting therapy.

As used herein the term "prognostic marker" shall be taken to mean an indicator of the likelihood of progression of the disease, in particular aggressiveness and metastatic potential of a breast tumour. It is preferably used to define patients with high, low and intermediate risks of death or recurrence after treatment that result from the inherent heterogeneity of the disease process.

Indicators of tumour aggressivness standard in the art include but are not limited to tumour stage, tumour grade, nodal status and survival.

As used herein the term "survival" shall be taken to include survival until mortality also known as overall survival (wherein said mortality may be either irrespective of cause or breast tumour related); "recurrence-free survival" (wherein the term recurrence shall include both localised and distant recurrence) ; metastasis free survival; disease free survival (wherein the term disease shall include breast cancer and diseases associated therewith). The length of said survival may be calculated by reference to a defined start point (e.g. time of diagnosis or start of treatment) and end point (e.g. death, recurrence or metastasis).

As used herein the term 'predictive marker' shall be taken to mean an indicator of response to therapy, said response is preferably defined according to patient survival. As defined herein the term predictive marker may in some situations fall within the remit of a herein described 'prognostic marker'. The two terms shall not be taken to be mutually exclusive.

Using the methods and nucleic acids described herein, statistically significant models of patient disease free survival or metastasis free survival or overall survival and/or disease progression can be developed and utilised to assist patients and clinicians in determining suitable treatment options to be included in the therapeutic regimen.

In one aspect the described method is to be used to assess the utility of therapeutic regimens comprising one or more treatments which is either an aggressive therapy such as chemotherapy or a treatment which targets the estrogen receptor pathway or is involved in estrogen metabolism, production or secretion as a therapy for patients suffering from a cell proliferative

disorder of the breast tissues. In particular this aspect of the method enables the physician to determine which treatments may be used in addition to or instead of said endocrine treatment.

In a further aspect the described method enables the characterisation of the cell proliferative disorder in terms of aggressiveness, thereby enabling the physician to recommend suitable treatments. Thus, the present invention will be seen to reduce the problems associated with present breast cell proliferative disorder treatment response prediction methods.

Using the methods and nucleic acids as described herein, patient survival can be evaluated before or during treatment for a cell proliferative disorder of the breast tissues, in order to provide critical information to the patient and clinician as to the likely progression of the disease. It will be appreciated, therefore, that the methods and nucleic acids exemplified herein can serve to improve a patient's quality of life and odds of treatment success by allowing both patient and clinician a more accurate assessment of the patient's treatment options.

The method according to the definition may be used for the improved treatment of all breast cell proliferative disorder patients, both pre and post menopausal and independent of their node or estrogen receptor status. However, it is particularly preferred that said patients are node-negative and estrogen receptor positive.

The present invention makes available a method for the improved treatment and monitoring of breast cell proliferative disorders, by enabling the accurate prediction of a patient's survival with endocrine therapy comprising one or more treatments which target the estrogen receptor pathway or are involved in estrogen metabolism, production, or secretion.

In a particularly preferred embodiment, the method according to the invention enables the differentiation between patients who have a high risk of relapse under said endocrine therapy and those who have a low risk of relapse under said therapy. The method enables the determination of a methylation pattern characteristic for a predicted survival time, in addition to the characterisation of tumours in terms of aggressiveness.

The method according to the invention may be used for the analysis of a wide variety of cell proliferative disorders of the breast tissues including, but not limited to, ductal carcinoma *in situ*, invasive ductal carcinoma, invasive lobular carcinoma, lobular carcinoma *in situ*, come-

docarcinoma, inflammatory carcinoma, mucinous carcinoma, scirrhous carcinoma, colloid carcinoma, tubular carcinoma, medullary carcinoma, metaplastic carcinoma, and papillary carcinoma and papillary carcinoma *in situ*, undifferentiated or anaplastic carcinoma and Paget's disease of the breast.

The method according to the invention is particularly suited to the prediction of survival of breast cancer in the following treatment setting. In one embodiment, the method is applied to patients who receive endocrine pathway targeting treatment as secondary treatment to an initial non chemotherapeutical therapy, e.g. surgery (hereinafter referred to as the adjuvant setting) as illustrated in Figure 1. Such a treatment is often prescribed to patients suffering from Stage 1 to 3 breast carcinomas. In this embodiment patients survival times are predicted according to their gene expression or genetic or epigenetic modifications. By detecting patients with worse disease free survival times the physician may choose to recommend the patient for further treatment, instead of or in addition to the endocrine targeting therapy(s), in particular but not limited to, chemotherapy.

The herein described invention provides a novel breast cell proliferative disorder prognostic biomarker. It is herein described that aberrant expression of the gene PITX2 is correlated to prognosis of breast cell proliferative disorder patients, in particular breast carcinoma. In particular this marker provides a novel means for the characterisation of breast carcinomas. As described herein determination of the expression of the gene PITX2 enables the prediction of survival (or outcome) of a patient treated with one or more treatments which target the estrogen receptor, synthesis or conversion pathways or are otherwise involved in estrogen metabolism, production or secretion. Survival or outcome may be based on the patient's survival or clinical or pathological tumour response, or response measured with other surrogate parameters.

The herein described invention is thereby useful for the differentiation of individuals who may be appropriately treated with one or more treatments which target the estrogen receptor pathway or are involved in estrogen metabolism, production or secretion from those individuals who would be optimally treated with other treatments in addition to said treatment. Preferred 'other treatments' include but are not limited to chemotherapy or radiotherapy.

In a further embodiment of the invention the aberrant expression of a plurality of genes comprising the gene PITX2 is analysed. Said plurality of genes is hereinafter referred to as a 'gene panel'. The analysis of multiple genes increases the accuracy of a prognosis. It is preferred that the gene panel consists of up to seven genes and/or their promoter regions associated with prognosis of breast carcinoma patients. It is further preferred that said panel consists of the gene PITX2 and one or more genes selected from the group consisting ABCA8, CDK6, ERBB2, ONECUT2, PLAU, TBC1D3 and TFF1. It is particularly preferred that the gene panel is selected from the group of gene panels consisting of:

- PITX2, PLAU & TFF1
- PITX2 & PLAU
- PITX2 & TFF1

This invention therefore relates to new methods and sequences for the evaluation of adjuvant therapy of patients diagnosed with breast cell proliferative disease based on a prediction of survival or outcome.

More specifically this invention provides new methods and sequences, for patients diagnosed with breast cell proliferative disease, allowing the evaluation of adjuvant therapy, i.e. therapy before or after surgical removal of the tumour, like a cytotoxic therapy (chemotherapy) in addition to or instead of (for example in hormone receptor negative patients) an endocrine treatment, like treatment with Tamoxifen or aromatase inhibitors, wherein the evaluation is based on the prediction of the patient's survival.

One aspect of the invention is the provision of tools for predicting the survival of a patient diagnosed with a breast cell proliferative disease, such as breast cancer. These tools comprise methods for the analysis of either the expression levels of PITX2 protein, or PITX2 mRNA or the analysis of the patient's individual genetic or epigenetic modification of the gene PITX2 – summarised as the analysis of expression of the gene PITX2. Preferably the invention relates to methods for predicting the survival of a patient diagnosed with breast cancer. Preferably said patient is treated with at least one adjuvant endocrine treatment, wherein endocrine treatment is meant to comprise any treatment targeting the estrogen receptor pathway or estrogen synthesis pathway or estrogen conversion pathway i.e., which is involved in estrogen metabolism, production or secretion. Preferably the patient was treated with said adjuvant

endocrine treatment after surgical removal of the tumour. Also preferably the survival is the disease free survival.

Especially preferred are methods applied for the prediction of the disease free survival of a patient diagnosed with breast cancer under adjuvant endocrine treatment after surgical tumour removal. Even more preferred are those methods, which analyse the DNA methylation profile of the genomic region associated with the gene PITX2. Especially preferred is the analysis of the DNA methylation profile of the genomic sequence given in SEQ ID NO: 1. Especially preferred is furthermore the analysis of the methylation status of eight specific CpG dinucleotides, covered in the three sub-sequences of said SEQ ID NO: 1 given in SEQ ID NO: 13, 18 and 19. The use of nucleic acids hybridising to these nucleic acid sequences for the prediction of survival according to the invention are preferred embodiments of said invention. The use of nucleic acids hybridising to CpG positions within these nucleic acid sequences after these nucleic acids have been contacted with one or more agents that convert cytosine bases that are unmethylated at the 5'-position thereof to a base that is detectably dissimilar to cytosine in terms of hybridisation properties, for the prediction of survival according to the invention are especially preferred embodiments of said invention.

This methodology presents further improvements over the state of the art in that the method may be applied to any subject, independent of the estrogen and/or progesterone receptor status. Therefore in a preferred embodiment, the subject is not required to have been tested for estrogen or progesterone receptor status.

The object of the invention is preferably achieved by means of the analysis of the methylation pattern of PITX2 and/or its regulatory region. In a particularly preferred embodiment the sequence of said gene comprises SEQ ID NO: 1 and the sequence complementary thereto.

In one preferred embodiment the object of the invention is the prediction of survival of a subject under a treatment which targets the estrogen receptor pathway or is involved in estrogen metabolism, production or secretion. This is achieved by analysis of the expression of PITX2 and wherein it is further preferred that the sequence of said gene comprises SEQ ID NO: 1 or parts thereof.

In one aspect the invention discloses novel methods utilising the gene PITX2 for the prediction of survival of a patient diagnosed with a breast cell proliferative disease. In a preferred embodiment said patient diagnosed with a breast cell proliferative disease is treated with adjuvant endocrine monotherapy.

The invention discloses the use of the gene PITX2, as well as its promoter and regulatory elements as a prognostic marker for survival of breast cancer patients. It is preferred that these patients are treated with adjuvant endocrine monotherapy. Furthermore, the disclosed method shows the applicability of said gene to answer the question and help guiding the decision whether or not an adjuvant chemotoxic therapy shall be prescribed, preferably in addition to endocrine treatment, like the treatment with tamoxifen or aromatase inhibitors.

In one aspect of the invention, the disclosed matter provides novel nucleic acid sequences useful for the analysis of methylation within said gene, other aspects provide novel uses of the gene and the gene product as well as methods, assays and kits directed to prognosing the survival of a patient diagnosed with breast cell proliferative disease. Preferably said patient is treated with adjuvant endocrine monotherapy.

In one embodiment the method discloses the use of the gene PITX2 as a marker for the prognosis of the survival of a patient suffering from a breast cell proliferative disease. Preferably said patient is treated with adjuvant endocrine monotherapy. Said use of the gene may be enabled by means of any analysis of the expression of the gene, by means of mRNA expression analysis or protein expression analysis or by analysis of its genetic modifications leading to an altered expression (including LOH). However, in the most preferred embodiment of the invention, prediction of the survival of a patient diagnosed with breast cell proliferative disease, preferably treated with adjuvant endocrine monotherapy, is enabled by means of analysis of the methylation status of CpG sites within the gene PITX2 and its promoter or regulatory elements.

In one embodiment of the method aberrant expression of the gene PITX2 may be detected by analysis of loss of heterozygosity of the gene. In a first step genomic DNA is isolated from a biological sample of the patient's tumour. The isolated DNA is then analysed for LOH by any means standard in the art including but not limited to amplification of the gene locus or associated microsatellite markers. Said amplification may be carried out by any means standard in

the art including polymerase chain reaction (PCR), strand displacement amplification (SDA)and isothermal amplification.

The level of amplificate is then detected by any means known in the art including but not limited to gel electrophoresis and detection by probes (including Real Time PCR). Furthermore the amplificates may be labelled in order to aid said detection. Suitable detectable labels include but are not limited to fluorescence label, radioactive labels and mass labels the suitable use of which shall be described herein.

The detection of a decreased amount of an amplificate corresponding to one of the amplified alleles in a test sample as relative to that of a heterozygous control sample is indicative of LOH.

To detect the levels of mRNA encoding PITX2 in a detection system for breast cancer relapse, a sample is obtained from a patient. Said obtaining of a sample is not meant to be retrieving of a sample, as in performing a biopsy, but rather directed to the availability of an isolated biological material representing a specific tissue, relevant for the intended use. The sample can be a tumour tissue sample from the surgically removed tumour, a biopsy sample as taken by a surgeon and provided to the analyst or a sample of blood, plasma, serum or the like. The sample may be treated to extract the nucleic acids contained therein. The resulting nucleic acid from the sample is subjected to gel electrophoresis or other separation techniques. Detection involves contacting the nucleic acids and in particular the mRNA of the sample with a DNA sequence serving as a probe to form hybrid duplexes. The stringency of hybridisation is determined by a number of factors during hybridisation and during the washing procedure, including temperature, ionic strength, length of time and concentration of formamide. These factors are outlined in, for example, Sambrook et al. (Molecular Cloning: A Laboratory Manual, 2nd ed., 1989). Detection of the resulting duplex is usually accomplished by the use of labelled probes. Alternatively, the probe may be unlabeled, but may be detectable by specific binding with a ligand which is labelled, either directly or indirectly. Suitable labels and methods for labelling probes and ligands are known in the art, and include, for example, radioactive labels which may be incorporated by known methods (e.g., nick translation or kinasing), biotin, fluorescent groups, chemiluminescent groups (e.g., dioxetanes, particularly triggered dioxetanes), enzymes, antibodies, and the like.

In order to increase the sensitivity of the detection in a sample of mRNA encoding PITX2, the technique of reverse transcription/polymerisation chain reaction can be used to amplify cDNA transcribed from mRNA encoding PITX2. The method of reverse transcription/PCR is well known in the art (for example, see Watson and Fleming, supra).

The reverse transcription/PCR method can be performed as follows. Total cellular RNA is isolated by, for example, the standard guanidium isothiocyanate method and the total RNA is reverse transcribed. The reverse transcription method involves synthesis of DNA on a template of RNA using a reverse transcriptase enzyme and a 3' end primer. Typically, the primer contains an oligo(dT) sequence. The cDNA thus produced is then amplified using the PCR method and PITX2 specific primers. (Belyavsky et al, Nucl Acid Res 17:2919-2932, 1989; Krug and Berger, Methods in Enzymology, Academic Press,N.Y., Vol.152, pp. 316-325, 1987 which are incorporated by reference)

The present invention may also be described in certain embodiments as a kit for use in predicting the survival of a breast cancer patient before or after surgical tumour removal with or without adjuvant endocrine monotherapy state through testing of a biological sample. A representative kit may comprise one or more nucleic acid segments as described above that selectively hybridise to PITX2 mRNA and a container for each of the one or more nucleic acid segments. In certain embodiments the nucleic acid segments may be combined in a single tube. In further embodiments, the nucleic acid segments may also include a pair of primers for amplifying the target mRNA. Such kits may also include any buffers, solutions, solvents, enzymes, nucleotides, or other components for hybridisation, amplification or detection reactions. Preferred kit components include reagents for reverse transcription-PCR, in situ hybridisation, Northern analysis and/or RPA.

The present invention further provides for methods to detect the presence of the polypeptide, PITX2, in a sample obtained from a patient. It is preferred that said sequence is essentially the same as the sequence presented in SEQ ID NO: 20, as given in figure 10. Any method known in the art for detecting proteins can be used. Such methods include, but are not limited to immunodiffusion, immunoelectrophoresis, immunochemical methods, binder-ligand assays, immunohistochemical techniques, agglutination and complement assays. (for example see Basic and Clinical Immunology, Sites and Terr, eds., Appleton & Lange, Norwalk, Conn. pp 217-262, 1991 which is incorporated by reference). Preferred are binder-ligand immunoassay

methods including reacting antibodies with an epitope or epitopes of PITX2 and competitively displacing a labelled PITX2 protein or derivative thereof.

Certain embodiments of the present invention comprise the use of antibodies specific to the polypeptide encoded by the PITX2 gene. Such antibodies may be useful for prognosing the survival of a breast cancer patient preferably under adjuvant endocrine monotherapy by comparing a patient's levels of PITX2 marker expression to expression of the same marker in normal individuals. In certain embodiments production of monoclonal or polyclonal antibodies can be induced by the use of the PITX2 polypeptide as antigene. Such antibodies may in turn be used to detect expressed proteins as markers for prognosis of relapse of a breast cancer patient under adjuvant endocrine monotherapy. The levels of such proteins present in the peripheral blood of a patient may be quantified by conventional methods. Antibody-protein binding may be detected and quantified by a variety of means known in the art, such as labelling with fluorescent or radioactive ligands. The invention further comprises kits for performing the above-mentioned procedures, wherein such kits contain antibodies specific for the PITX2 polypeptides.

Numerous competitive and non-competitive protein binding immunoassays are well known in the art. Antibodies employed in such assays may be unlabeled, for example as used in agglutination tests, or labelled for use a wide variety of assay methods. Labels that can be used include radionuclides, enzymes, fluorescers, chemiluminescers, enzyme substrates or co-factors, enzyme inhibitors, particles, dyes and the like for use in radioimmunoassay (RIA), enzyme immunoassays, e.g., enzyme-linked immunosorbent assay (ELISA), fluorescent immunoassays and the like. Polyclonal or monoclonal antibodies to PITX2 or an epitope thereof can be made for use in immunoassays by any of a number of methods known in the art. One approach for preparing antibodies to a protein is the selection and preparation of an amino acid sequence of all or part of the protein, chemically synthesising the sequence and injecting it into an appropriate animal, usually a rabbit or a mouse (Milstein and Kohler Nature 256:495-497, 1975; Gulfre and Milstein, Methods in Enzymology: Immunochemical Techniques 73:1-46, Langone and Banatis eds., Academic Press, 1981 which are incorporated by reference). Methods for preparation of PITX2 or an epitope thereof include, but are not limited to chemical synthesis, recombinant DNA techniques or isolation from biological samples.

The invention provides significant improvements over the state of the art in that -at the time of filing- there are no single markers known to the public which can be used to predict the likelihood of relapse or of survival of a breast cancer patient under adjuvant endocrine monotherapy, neither from tissue samples nor from body fluid samples.

Also, no methylation marker is known which can be used to detect the likelihood of relapse or of survival of a breast cancer patient. Especially, no methylation marker is known which can be used to detect the likelihood of relapse or of survival of a breast cancer patient under adjuvant endocrine monotherapy, neither from tissue samples nor from body fluid samples.

The objective of the invention is also preferably achieved by analysis of the methylation state of the CpG dinucleotides within the genomic sequence according to SEQ ID NO: 1 and sequences complementary thereto. SEQ ID NO: 1 discloses the gene PITX2 and its promoter and regulatory elements, wherein said fragment comprises CpG dinucleotides exhibiting a disease specific methylation pattern. The methylation pattern of the gene PITX2 and its promoter and regulatory elements have heretofore not been analysed with regard to prognosis or prediction of survival of a patient diagnosed with a breast cell proliferative disorder. Due to the degeneracy of the genetic code, the sequence as identified in SEQ ID NO: 1 should be interpreted so as to include all substantially similar and equivalent sequences upstream of the promoter region of a gene which encodes a polypeptide with the biological activity of that encoded by PITX2.

In a preferred embodiment of the method, the objective of the invention is achieved by analysis of a nucleic acid comprising a sequence of at least 18 bases in length according to one of SEQ ID NO: 2 to SEQ ID NO: 5 and sequences complementary thereto.

The sequences of SEQ ID NO: 2 to 5 provide modified versions of the nucleic acid according to SEQ ID NO: 1, wherein the conversion of said sequence results in the synthesis of a nucleic acid having a sequence that is unique and distinct from SEQ ID NO: 1 as follows. (*see also* the following TABLE 1): SEQ ID NO: 1, sense DNA strand of PITX2 gene and its promoter and regulatory elements; SEQ ID NO: 2, converted SEQ ID NO: 1, wherein "C" converted to "T," but "CpG" remains "CpG." (*i.e.*, corresponds to case where, for SEQ ID NO: 1, all "C" residues of CpG dinucleotide sequences are methylated and are thus not converted); SEQ ID NO: 3, complement of SEQ ID NO: 1, wherein "C" converted to "T," but "CpG"

remains "CpG" (*i.e.*, corresponds to case where, for the complement (*anti*sense strand) of SEQ ID NO: 1, all "C" residues of CpG dinucleotide sequences are methylated and are thus not converted); SEQ ID NO: 4, converted SEQ ID NO: 1, wherein "C" converted to "T" for all "C" residues, including those of "CpG" dinucleotide sequences (*i.e.*, corresponds to case where, for SEQ ID NO: 1, all "C" residues of CG dinucleotide sequences are unmethylated); SEQ ID NO: 5, complement of SEQ ID NO: 1, wherein "C" converted to "T" for all "C" residues, including those of "CpG" dinucleotide sequences (*i.e.*, corresponds to case where, for the complement (*anti*sense strand) of SEQ ID NO: 1, all "C" residues of CpG dinucleotide sequences are unmethylated).

TABLE 1. Description of SEQ ID NO: 1 to 5

SEQ ID NO: NO	**Relationship to SEQ ID NO:1**	**Nature of cytosine base conversion**
SEQ ID NO:1	Sense strand (PITX2 gene including promoter and regulatory elements)	None; untreated sequence
SEQ ID NO:2	Converted sense strand	"C" to" T," but "CpG" remains "CpG" (all "C" residues of CpGs are methylated)
SEQ ID NO:3	Converted antisense strand	"C" to"T," but "CpG" remains "CpG" (all "C" residues of CpGs are methylated)
SEQ ID NO:4	Converted sense strand	"C" to"T" for all "C" residues (all "C" residues of CpGs are unmethylated)
SEQ ID NO:5	Converted antisense strand	"C" to "T" for all "C" residues (all "C" residues of CpGs are unmethylated)

Significantly, heretofore, the nucleic acid sequences and molecules according to SEQ ID NO: 1 to SEQ ID NO: 5 were not implicated in or connected with the ascertainment of the prognosis of breast cancer relapse or the prediction of survival of breast cancer patients.

The described invention further discloses oligonucleotides or oligomers for detecting the cytosine methylation state within pretreated DNA, according to SEQ ID NO: 2 to SEQ ID NO: 5. The use of said oligonucleotides or oligomers comprising a nucleic acid sequence having a length of at least nine (9) nucleotides which hybridise, under moderately stringent or stringent

conditions (as defined herein above), to a pretreated nucleic acid sequence according to SEQ ID NO: 2 to SEQ ID NO: 5 and/or sequences complementary thereto is another embodiment of this invention.

Thus, the present invention includes the use of nucleic acid molecules (*e.g.*, oligonucleotides and peptide nucleic acid (PNA) molecules (PNA-oligomers)) that hybridise under moderately stringent and/or stringent hybridisation conditions to all or a portion of the sequences of SEQ ID NO: 2 to 5, or to the complements thereof for prediction of survival according to the invention. The hybridising portion of the hybridising nucleic acids is typically at least 9, 15, 20, 25, 30 or 35 nucleotides in length. However, longer molecules have inventive utility, and are thus within the scope of the present invention.

Preferably, the hybridising portion of the inventive hybridising nucleic acids is at least 95%, or at least 98%, or 100% identical to the sequence, or to a portion thereof of SEQ ID NO: 2 to 5, or to the complements thereof.

Hybridising nucleic acids of the type described herein can be used, for example, as a primer (*e.g.*, a PCR primer), or a diagnostic and/or prognostic probe or primer. Preferably, hybridisation of the oligonucleotide probe to a nucleic acid sample is performed under stringent conditions and the probe is 100% identical to the target sequence. Nucleic acid duplex or hybrid stability is expressed as the melting temperature or Tm, which is the temperature at which a probe dissociates from a target DNA. This melting temperature is used to define the required stringency conditions.

For target sequences that are related and substantially identical to the corresponding sequence of SEQ ID NO: 1 (such as PITX2 allelic variants and SNPs), rather than identical, it is useful to first establish the lowest temperature at which only homologous hybridisation occurs with a particular concentration of salt (*e.g.*, SSC or SSPE). Then, assuming that 1% mismatching results in a 1°C decrease in the Tm, the temperature of the final wash in the hybridisation reaction is reduced accordingly (for example, if sequences having > 95% identity with the probe are sought, the final wash temperature is decreased by 5°C). In practice, the change in Tm can be between 0.5°C and 1.5°C per 1% mismatch.

Examples of inventive oligonucleotides of length X (in nucleotides), as indicated by polynu-

cleotide positions with reference to, *e.g.*, SEQ ID NO: 1, include those corresponding to sets of consecutively overlapping oligonucleotides of length X, where the oligonucleotides within each consecutively overlapping set (corresponding to a given X value) are defined as the finite set of Z oligonucleotides from nucleotide positions:

n to (n + (X-1));

where n=1, 2, 3,...(Y-(X-1));

where Y equals the length (nucleotides or base pairs) of SEQ ID NO: 1 ;

where X equals the common length (in nucleotides) of each oligonucleotide in the set (*e.g.*, X=20 for a set of consecutively overlapping 20-mers); and

where the number (Z) of consecutively overlapping oligomers of length X for a given SEQ ID NO: NO of length Y is equal to Y-(X-1). For example Z=2,785-19=2,766 for either sense or antisense sets of SEQ ID NO: 1, where X=20.

Preferably, the set is limited to those oligomers that comprise at least one CpG, Cpa or tpG dinucleotide, wherein 'Cpa' is indicating that said Cpa hybridises to a position (tpG) which was a CpG prior to bisulfite conversion and is a TpG now; and wherein 'tpG' is indicating that said tpG hybridises to a position (Cpa) which is the complementary to a position (tpG) which was a CpG prior to bisulfite conversion and is a TpG now.

The present invention encompasses, for *each* of SEQ ID NO: 2 to 5 (sense and antisense), the use of multiple consecutively overlapping sets of oligonucleotides or modified oligonucleotides of length X, where, *e.g.*, X= 9, 10, 17, 20, 22, 23, 25, 27, 30 or 35 nucleotides.

The oligonucleotides or oligomers according to the present invention constitute effective tools useful to ascertain genetic and epigenetic parameters of the genomic sequence corresponding to SEQ ID NO: 1. Preferred sets of such oligonucleotides or modified oligonucleotides of length X are those consecutively overlapping sets of oligomers corresponding to SEQ ID NO:1-5 (and to the complements thereof). Preferably, said oligomers comprise at least one CpG, tpG or Cpa dinucleotide.

Particularly preferred oligonucleotides or oligomers used to the present invention are those in which the cytosine of the CpG dinucleotide (or of the corresponding converted TpG or CpA dinucleotide) sequences is within the middle third of the oligonucleotide; that is, where the

oligonucleotide is, for example, 13 bases in length, the CpG, TpG or CpA dinucleotide is positioned within the fifth to ninth nucleotide from the 5'-end.

The oligonucleotides used in this invention can also be modified by chemically linking the oligonucleotide to one or more moieties or conjugates to enhance the activity, stability or detection of the oligonucleotide. Such moieties or conjugates include chromophores, fluorophors, lipids such as cholesterol, cholic acid, thioether, aliphatic chains, phospholipids, polyamines, polyethylene glycol (PEG), palmityl moieties, and others as disclosed in, for example, United States Patent Numbers 5,514,758, 5,565,552, 5,567,810, 5,574,142, 5,585,481, 5,587,371, 5,597,696 and 5,958,773. The probes may also exist in the form of a PNA (peptide nucleic acid) which has particularly preferred pairing properties. Thus, the oligonucleotide may include other appended groups such as peptides, and may include hybridisation-triggered cleavage agents (Krol et al., *BioTechniques* 6:958-976, 1988) or intercalating agents (Zon, *Pharm. Res*. 5:539-549, 1988). To this end, the oligonucleotide may be conjugated to another molecule, *e.g.*, a chromophore, fluorophor, peptide, hybridisation-triggered cross-linking agent, transport agent, hybridisation-triggered cleavage agent, etc.

The oligonucleotide may also comprise at least one art-recognised modified sugar and/or base moiety, or may comprise a modified backbone or non-natural internucleoside linkage.

The oligomers used in the present invention are normally used in so called "sets" which contain at least one oligomer for analysis of each of the CpG dinucleotides of a genomic sequence comprising SEQ ID NO: 1 and sequences complementary thereto or to their corresponding CG, tG or Ca dinucleotide within the pretreated nucleic acids according to SEQ ID NO: 2 to SEQ ID NO: 5 and sequences complementary thereto, wherein a 't' indicates a nucleotide which converted from a cytosine into a thymine and wherein 'a' indicates the complementary nucleotide to such a converted thymine. Preferred is a set which contains at least one oligomer for each of the CpG dinucleotides within the gene PITX2 and it's promoter and regulatory elements in both the pretreated and genomic versions of said gene, SEQ ID NO: 2 to 5 and SEQ ID NO: 1, respectively. However, it is anticipated that for economic or other factors it may be preferable to analyse a limited selection of the CpG dinucleotides within said sequences and the contents of the set of oligonucleotides should be altered accordingly. Therefore, the present invention moreover relates to a set of at least 3 n (oligonucleotides and/or PNA-oligomers) used for detecting the cytosine methylation state in pretreated geno-

mic DNA (SEQ ID NO: 2 to SEQ ID NO: 5 and sequences complementary thereto) and genomic DNA (SEQ ID NO: 1 and sequences complementary thereto). These probes enable diagnosis and/or therapy of genetic and epigenetic parameters of cell proliferative disorders. The set of oligomers may also be used for detecting single nucleotide polymorphisms (SNPs) in pretreated genomic DNA (SEQ ID NO: 2 to SEQ ID NO: 5, and sequences complementary thereto) and genomic DNA (SEQ ID NO: 1, and sequences complementary thereto) .

Moreover, the present invention includes the use of a set of at least two oligonucleotides which can be used as so-called "primer oligonucleotides" for amplifying DNA sequences of one of SEQ ID NO: 1 to SEQ ID NO: 5 and sequences complementary thereto, or segments thereof.

In the case of the sets of oligonucleotides according to the present invention, it is preferred that at least one and more preferably all members of the set of oligonucleotides is bound to a solid phase.

According to the present invention, it is preferred that an arrangement of different oligonucleotides and/or PNA-oligomers (a so-called "array") made available by the present invention is present in a manner that it is likewise bound to a solid phase. This array of different oligonucleotide- and/or PNA-oligomer sequences can be characterised in that it is arranged on the solid phase in the form of a rectangular or hexagonal lattice. The solid phase surface is preferably composed of silicon, glass, polystyrene, aluminium, steel, iron, copper, nickel, silver, or gold. However, nitrocellulose as well as plastics such as nylon which can exist in the form of pellets or also as resin matrices may also be used.

Therefore, a further subject matter of the present invention is a method for manufacturing an array fixed to a carrier material for analysis in connection with cell proliferative disorders, in which method at least one oligomer according to the present invention is coupled to a solid phase. Methods for manufacturing such arrays are known, for example, from US Patent 5,744,305 by means of solid-phase chemistry and photolabile protecting groups.

A further subject matter of the present invention relates to a DNA chip for the analysis of cell proliferative disorders. DNA chips are known, for example, in US Patent 5,837,832.

The described invention further provides a composition of matter useful for prognosing the relapse of breast cancer patients or predicting the survival of breast cancer patients. Said composition comprising at least one nucleic acid 18 base pairs in length of a segment of the nucleic acid sequence disclosed in SEQ ID NO: 2 to 5, and one or more substances taken from the group comprising :

1-5 mM Magnesium Chloride, 100-500 µM dNTP, 0.5-5 units/10µl of taq polymerase, bovine serum albumen, an oligomer in particular an oligonucleotide or peptide nucleic acid (PNA)-oligomer, said oligomer comprising in each case at least one base sequence having a length of at least 9 nucleotides which is complementary to, or hybridises under moderately stringent or stringent conditions to a pretreated genomic DNA according to one of the SEQ ID NO: 2 to SEQ ID NO: 5 and sequences complementary thereto. It is preferred that said composition of matter comprises a buffer solution appropriate for the stabilisation of said nucleic acid in an aqueous solution and enabling polymerase based reactions within said solution.. Suitable buffers are known in the art and commercially available.

The present invention further provides a method for conducting an assay in order to ascertain genetic and/or epigenetic parameters of the gene PITX2 and its promoter and regulatory elements. Most preferably the assay according to the following method is used in order to detect methylation within the gene PITX2 wherein said methylated nucleic acids are present in a solution further comprising an excess of background DNA, wherein the background DNA is present in between 100 to 1000 times the concentration of the DNA to be detected. Said method comprising contacting a nucleic acid sample obtained from said subject with at least one reagent or a series of reagents, wherein said reagent or series of reagents, distinguishes between methylated and non-methylated CpG dinucleotides within the target nucleic acid.

Preferably, said method comprises the following steps: In the first *step*, a sample of the tissue to be analysed is obtained. The source may be any suitable source, preferably, the source of the sample is selected from the group consisting of histological slides, biopsies, paraffin-embedded tissue, bodily fluids, plasma, serum, stool, urine, blood, nipple aspirate and combinations thereof. Preferably, the source is tumour tissue, biopsies, serum, urine, blood or nipple aspirate. The most preferred source, is the tumour sample, surgically removed from the patient or a biopsy sample of said patient.

The DNA is then isolated from the sample. Extraction may be by means that are standard to

one skilled in the art, including the use of detergent lysates, sonification and vortexing with glass beads. Once the nucleic acids have been extracted, the genomic double stranded DNA is used in the analysis.

In the second step of the method, the genomic DNA sample is treated in such a manner that cytosine bases which are unmethylated at the 5'-position are converted to uracil, thymine, or another base which is dissimilar to cytosine in terms of hybridisation behaviour. This will be understood as 'pretreatment' herein.

The above described pretreatment of genomic DNA is preferably carried out with bisulfite (hydrogen sulfite, disulfite) and subsequent alkaline hydrolysis which results in a conversion of non-methylated cytosine nucleobases to uracil or to another base which is dissimilar to cytosine in terms of base pairing behaviour. Enclosing the DNA to be analysed in an agarose matrix, thereby preventing the diffusion and renaturation of the DNA (bisulfite only reacts with single-stranded DNA), and replacing all precipitation and purification steps with fast dialysis (Olek A, et al., A modified and improved method for bisulfite based cytosine methylation analysis, *Nucleic Acids Res.* 24:5064-6, 1996) is one preferred example how to perform said pretreatment. It is further preferred that the bisulfite treatment is carried out in the presence of a radical scavenger or DNA denaturing agent.

In the third step of the method, fragments of the pretreated DNA are amplified. Wherein the source of the DNA is free DNA from serum, or DNA extracted from paraffin it is particularly preferred that the size of the amplificate fragment is between 100 and 200 base pairs in length, and wherein said DNA source is extracted from cellular sources (e.g. tissues, biopsies, cell lines) it is preferred that the amplificate is between 100 and 350 base pairs in length. It is particularly preferred that said amplificates comprise at least one 20 base pair sequence comprising at least three CpG dinucleotides. Said amplification is carried out using sets of primer oligonucleotides according to the present invention, and a preferably heat-stable polymerase. The amplification of several DNA segments can be carried out simultaneously in one and the same reaction vessel, in one embodiment of the method preferably six or more fragments are amplified simultaneously. Typically, the amplification is carried out using a polymerase chain reaction (PCR). The set of primer oligonucleotides includes at least two oligonucleotides whose sequences are each reverse complementary, identical, or hybridise under stringent or highly stringent conditions to an at least 18-base-pair long segment of the base sequences of

SEQ ID NO: 2 to SEQ ID NO: 5 and sequences complementary thereto.

In one especially preferred embodiment of the method the primers may be selected form the group consisting to SEQ ID NO: 6 to SEQ ID NO: 10.

In an alternate embodiment of the method, the methylation status of preselected CpG positions within the nucleic acid sequences comprising SEQ ID NO: 2 to SEQ ID NO: 5 may be detected by use of methylation-specific primer oligonucleotides. This technique (MSP) has been described in United States Patent No. 6,265,171 to Herman. The use of methylation status specific primers for the amplification of bisulfite treated DNA allows the differentiation between methylated and unmethylated nucleic acids. MSP primers pairs contain at least one primer which hybridises to a bisulfite treated CpG dinucleotide. Therefore, the sequence of said primers comprises at least one CpG , TpG or CpA dinucleotide. MSP primers specific for non-methylated DNA contain a "T' at the 3' position of the C position in the CpG. Preferably, therefore, the base sequence of said primers is required to comprise a sequence having a length of at least 18 nucleotides which hybridises to a pretreated nucleic acid sequence according to SEQ ID NO: 2 to SEQ ID NO: 5 and sequences complementary thereto, wherein the base sequence of said oligomers comprises at least one CpG, tpG or Cpa dinucleotide. In this embodiment of the method according to the invention it is particularly preferred that the MSP primers comprise between 2 and 4 CpG , tpG or Cpa dinucleotides. It is further preferred that said dinucleotides are located within the 3' half of the primer e.g. wherein a primer is 18 bases in length the specified dinucleotides are located within the first 9 bases form the 3'end of the molecule. In addition to the CpG , tpG or Cpa dinucleotides it is further preferred that said primers should further comprise several bisulfite converted bases (i.e. cytosine converted to thymine, or on the hybridising strand, guanine converted to adenosine). In a further preferred embodiment said primers are designed so as to comprise no more than 2 cytosine or guanine bases.

The fragments obtained by means of the amplification can carry a directly or indirectly detectable label. Preferred are labels in the form of fluorescence labels, radionuclides, or detachable molecule fragments having a typical mass which can be detected in a mass spectrometer. Where said labels are mass labels, it is preferred that the labelled amplificates have a single positive or negative net charge, allowing for better detectability in the mass spectrometer. The detection may be carried out and visualised by means of, *e.g.*, matrix assisted laser desorp-

tion/ionisation mass spectrometry (MALDI) or using electron spray mass spectrometry (ESI).

Matrix Assisted Laser Desorption/Ionization Mass Spectrometry (MALDI-TOF) is a very efficient development for the analysis of biomolecules (Karas & Hillenkamp, *Anal Chem.*, 60:2299-301, 1988). An analyte is embedded in a light-absorbing matrix. The matrix is evaporated by a short laser pulse thus transporting the analyte molecule into the vapour phase in an unfragmented manner. The analyte is ionised by collisions with matrix molecules. An applied voltage accelerates the ions into a field-free flight tube. Due to their different masses, the ions are accelerated at different rates. Smaller ions reach the detector sooner than bigger ones. MALDI-TOF spectrometry is well suited to the analysis of peptides and proteins. The analysis of nucleic acids is somewhat more difficult (Gut & Beck, *Current Innovations and Future Trends*, 1:147-57, 1995). The sensitivity with respect to nucleic acid analysis is approximately 100-times less than for peptides, and decreases disproportionally with increasing fragment size. Moreover, for nucleic acids having a multiply negatively charged backbone, the ionisation process via the matrix is considerably less efficient. In MALDI-TOF spectrometry, the selection of the matrix plays an eminently important role. For the desorption of peptides, several very efficient matrixes have been found which produce a very fine crystallisation. There are now several responsive matrixes for DNA, however, the difference in sensitivity between peptides and nucleic acids has not been reduced. This difference in sensitivity can be reduced, however, by chemically modifying the DNA in such a manner that it becomes more similar to a peptide. For example, phosphorothioate nucleic acids, in which the usual phosphates of the backbone are substituted with thiophosphates, can be converted into a charge-neutral DNA using simple alkylation chemistry (Gut & Beck, *Nucleic Acids Res.* 23: 1367-73, 1995). The coupling of a charge tag to this modified DNA results in an increase in MALDI-TOF sensitivity to the same level as that found for peptides. A further advantage of charge tagging is the increased stability of the analysis against impurities, which makes the detection of unmodified substrates considerably more difficult.

In a particularly preferred embodiment of the method the amplification of step three is carried out in the presence of at least one species of blocker oligonucleotides. The use of such blocker oligonucleotides has been described by Yu et al., *BioTechniques* 23:714-720, 1997. The use of blocking oligonucleotides enables the improved specificity of the amplification of a subpopulation of nucleic acids. Blocking probes hybridised to a nucleic acid suppress, or hinder the polymerase mediated amplification of said nucleic acid. In one embodiment of the method

blocking oligonucleotides are designed so as to hybridise to background DNA. In a further embodiment of the method said oligonucleotides are designed so as to hinder or suppress the amplification of unmethylated nucleic acids as opposed to methylated nucleic acids or vice versa.

Blocking probe oligonucleotides are hybridised to the bisulfite treated nucleic acid concurrently with the PCR primers. PCR amplification of the nucleic acid is terminated at the 5' position of the blocking probe, such that amplification of a nucleic acid is suppressed where the complementary sequence to the blocking probe is present. The probes may be designed to hybridise to the bisulfite treated nucleic acid in a methylation status specific manner. For example, for detection of methylated nucleic acids within a population of unmethylated nucleic acids, suppression of the amplification of nucleic acids which are unmethylated at the position in question would be carried out by the use of blocking probes comprising a 'TpG' at the position in question, as opposed to a 'CpG.' In one embodiment of the method the sequence of said blocking oligonucleotides should be identical or complementary to molecule is complementary or identical to a sequence at least 18 base pairs in length selected from the group consisting of SEQ ID NO: 2 to 5, preferably comprising one or more CpG, TpG or CpA dinucleotides. In one embodiment of the method the sequence of said oligonucleotides is selected from the group consisting SEQ ID NO: 15 and SEQ ID NO: 16 and sequences complementary thereto.

For PCR methods using blocker oligonucleotides, efficient disruption of polymerase-mediated amplification requires that blocker oligonucleotides not be elongated by the polymerase. Preferably, this is achieved through the use of blockers that are 3'-deoxyoligonucleotides, or oligonucleotides derivatised at the 3' position with other than a "free" hydroxyl group. For example, 3'-O-acetyl oligonucleotides are representative of a preferred class of blocker molecule.

Additionally, polymerase-mediated decomposition of the blocker oligonucleotides should be precluded. Preferably, such preclusion comprises either use of a polymerase lacking 5'-3' exonuclease activity, or use of modified blocker oligonucleotides having, for example, thioate bridges at the 5'-termini thereof that render the blocker molecule nuclease-resistant. Particular applications may not require such 5' modifications of the blocker. For example, if the blocker- and primer-binding sites overlap, thereby precluding binding of the primer (*e.g.*, with

excess blocker), degradation of the blocker oligonucleotide will be substantially precluded. This is because the polymerase will not extend the primer toward, and through (in the 5'-3' direction) the blocker - a process that normally results in degradation of the hybridised blocker oligonucleotide.

A particularly preferred blocker/PCR embodiment, for purposes of the present invention and as implemented herein, comprises the use of peptide nucleic acid (PNA) oligomers as blocking oligonucleotides. Such PNA blocker oligomers are ideally suited, because they are neither decomposed nor extended by the polymerase.

In one embodiment of the method, the binding site of the blocking oligonucleotide is identical to, or overlaps with that of the primer and thereby hinders the hybridisation of the primer to its binding site. In a further preferred embodiment of the method, two or more such blocking oligonucleotides are used. In a particularly preferred embodiment, the hybridisation of one of the blocking oligonucleotides hinders the hybridisation of a forward primer, and the hybridisation of another of the probe (blocker) oligonucleotides hinders the hybridisation of a reverse primer that binds to the amplificate product of said forward primer.

In an alternative embodiment of the method, the blocking oligonucleotide hybridises to a location between the reverse and forward primer positions of the treated background DNA, thereby hindering the elongation of the primer oligonucleotides.

It is particularly preferred that the blocking oligonucleotides are present in at least 5 times the concentration of the primers.

In the fourth step of the method, the amplificates obtained during the third step of the method are analysed in order to ascertain the methylation status of the CpG dinucleotides prior to the treatment.

In embodiments where the amplificates were obtained by means of MSP amplification and/or blocking oligonucleotides, the presence or absence of an amplificate is in itself indicative of the methylation state of the CpG positions covered by the primers and or blocking oligonucleotide, according to the base sequences thereof.. All possible known molecular biological methods may be used for this detection, including, but not limited to gel electrophoresis, se-

quencing, liquid chromatography, hybridisations, real time PCR analysis or combinations thereof. This step of the method further acts as a qualitative control of the preceding steps.

In the fourth step of the method amplificates obtained by means of both standard and methylation specific PCR are further analysed in order to determine the CpG methylation status of the genomic DNA isolated in the first step of the method. This may be carried out by means of hybridisation-based methods such as, but not limited to, array technology and probe based technologies as well as by means of techniques such as sequencing and template directed extension.

In one embodiment of the method, the amplificates synthesised in step three are subsequently hybridised to an array or a set of oligonucleotides and/or PNA probes. In this context, the hybridisation takes place in the following manner: the set of probes used during the hybridisation is preferably composed of at least 2 oligonucleotides or PNA-oligomers; in the process, the amplificates serve as probes which hybridise to oligonucleotides previously bonded to a solid phase; the non-hybridised fragments are subsequently removed; said oligonucleotides contain at least one base sequence having a length of at least 9 nucleotides which is reverse complementary or identical to a segment of the base sequences specified in the SEQ ID NO: 2 to SEQ ID NO: 5; and the segment comprises at least one CpG , TpG or CpA dinucleotide.

In a preferred embodiment, said dinucleotide is present in the central third of the oligomer. For example, wherein the oligomer comprises one CpG dinucleotide, said dinucleotide is preferably the fifth to ninth nucleotide from the 5'-end of a 13-mer. One oligonucleotide exists for the analysis of each CpG dinucleotide within the sequence according to SEQ ID NO: 1, and the equivalent positions within SEQ ID NO: 2 to 5. Said oligonucleotides may also be present in the form of peptide nucleic acids. The non-hybridised amplificates are then removed. The hybridised amplificates are detected. In this context, it is preferred that labels attached to the amplificates are identifiable at each position of the solid phase at which an oligonucleotide sequence is located.

In yet a further embodiment of the method, the genomic methylation status of the CpG positions may be ascertained by means of oligonucleotide probes that are hybridised to the bisulfite treated DNA concurrently with the PCR amplification primers (wherein said primers may either be methylation specific or standard).

A particularly preferred embodiment of this method is the use of fluorescence-based Real Time Quantitative PCR (Heid et al., *Genome Res.* 6:986-994, 1996; *also see* United States Patent No. 6,331,393). There are two preferred embodiments of utilising this method. One embodiment, known as the TaqMan™ assay employs a dual-labelled fluorescent oligonucleotide probe. The TaqMan™ PCR reaction employs the use of a non-extendible interrogating oligonucleotide, called a TaqMan™ probe, which is designed to hybridise to a CpG-rich sequence located between the forward and reverse amplification primers. The TaqMan™ probe further comprises a fluorescent "reporter moiety" and a "quencher moiety" covalently bound to linker moieties (*e.g.*, phosphoramidites) attached to the nucleotides of the TaqMan™ oligonucleotide. Hybridised probes are displaced and broken down by the polymerase of the amplification reaction thereby leading to an increase in fluorescence. For analysis of methylation within nucleic acids subsequent to bisulfite treatment, it is required that the probe be methylation specific, as described in United States Patent No. 6,331,393, (hereby incorporated by reference in its entirety) also known as the MethyLight assay. The second preferred embodiment of this MethyLight technology is the use of dual-probe technology (Lightcycler®), each probe carrying donor or recipient fluorescent moieties, hybridisation of two probes in proximity to each other is indicated by an increase or fluorescent amplification primers. Both these techniques may be adapted in a manner suitable for use with bisulfite treated DNA, and moreover for methylation analysis within CpG dinucleotides.

Also any combination of these probes or combinations of these probes with other known probes may be used.

In a further preferred embodiment of the method, the fourth step of the method comprises the use of template-directed oligonucleotide extension, such as MS-SNuPE as described by Gonzalgo & Jones, *Nucleic Acids Res.* 25:2529-2531, 1997. In said embodiment it is preferred that the methylation specific single nucleotide extension primer (MS-SNuPE primer) is identical or complementary to a sequence at least nine but preferably no more than twenty five nucleotides in length of one or more of the sequences taken from the group of SEQ ID NO: 2 to SEQ ID NO: 5. However it is preferred to use fluorescently labelled nucleotides, instead of radiolabelled nucleotides.

In yet a further embodiment of the method, the fourth step of the method comprises sequenc-

ing and subsequent sequence analysis of the amplificate generated in the third step of the method (Sanger F., et al., *Proc Natl Acad Sci USA* 74:5463-5467, 1977).

Additional embodiments of the invention provide a method for the analysis of the methylation status of genomic DNA according to the invention (SEQ ID NO: 1) without the need for pre-treatment.

In the *first step* of such additional embodiments, the genomic DNA sample is isolated from tissue or cellular sources. Preferably, such sources include cell lines, histological slides, biopsy tissue, body fluids, or breast tumour tissue embedded in paraffin. Extraction may be by means that are standard to one skilled in the art, including but not limited to the use of detergent lysates, sonification and vortexing with glass beads. Once the nucleic acids have been extracted, the genomic double-stranded DNA is used in the analysis.

In a preferred embodiment, the DNA may be cleaved prior to the treatment, and this may be by any means standard in the state of the art, but preferably with methylation-sensitive restriction endonucleases.

In the *second step*, the DNA is then digested with one or more methylation sensitive restriction enzymes. The digestion is carried out such that hydrolysis of the DNA at the restriction site is informative of the methylation status of a specific CpG dinucleotide.

In the *third step*, which is optional but a preferred embodiment, the restriction fragments are amplified. This is preferably carried out using a polymerase chain reaction, and said amplificates may carry suitable detectable labels as discussed above, namely fluorophore labels, radionuclides and mass labels.

In the *final step* the amplificates are detected. The detection may be by any means standard in the art, for example, but not limited to, gel electrophoresis analysis, hybridisation analysis, incorporation of detectable tags within the PCR products, DNA array analysis, MALDI or ESI analysis.

The present invention enables prognosis of events which are disadvantageous to patients or individuals in which important genetic and/or epigenetic parameters within the PITX2 gene

and its promoter or regulatory elements may be used as prognostic markers for breast cancer relapse or as 'adjuvant marker' for prediction of need of additional treatment besides of endocrine monotherapy. Said parameters obtained by means of the present invention may be compared to another set of genetic and/or epigenetic parameters, the differences serving as the basis for a prognosis of events which are disadvantageous to patients or individuals.

Specifically, the present invention provides for prognostic cancer relapse assays based on measurement of differential methylation of PITX2 CpG dinucleotide sequences. Preferred gene sequences useful to measure such differential methylation are represented herein by SEQ ID NO: 1 to 5. Typically, such assays involve obtaining a tissue sample from a test tissue, performing an assay to measure the methylation status of at least one of the inventive PITX2-specific CpG dinucleotide sequences derived from the tissue sample, relative to a control sample, and making a diagnosis or prognosis or prediction based thereon.

In particular preferred embodiments, inventive oligomers to assess PITX2 specific CpG dinucleotide methylation status, such as those based on SEQ ID NO: 1 to 5, or arrays thereof, as well as a kit based thereon are used for the prognosis of breast cancer relapse and/or the prediction of survival of a patient diagnosed with breast cancer, preferably under endocrine treatment since surgical removal of the tumour.

Moreover, an additional aspect of the present invention is a kit comprising, for example: a bisulfite-containing reagent as well as at least one oligonucleotide whose sequences in each case correspond, are complementary, or hybridise under stringent or highly stringent conditions to a 18-base long segment of the sequences SEQ ID NO: 1 to 5. Said kit may further comprise instructions for carrying out and evaluating the described method. In a further preferred embodiment, said kit may further comprise standard reagents for performing a CpG position-specific methylation analysis, wherein said analysis comprises one or more of the following techniques: MS-SNuPE, MSP, MethyLight ™, HeavyMethyl™, COBRA, and nucleic acid sequencing. However, a kit along the lines of the present invention can also contain only part of the aforementioned components.

Typical reagents (*e.g.*, as might be found in a typical COBRA-based kit) for COBRA analysis may include, but are not limited to: PCR primers for specific gene (or methylation-altered DNA sequence or CpG island); restriction enzyme and appropriate buffer; gene-hybridisation

oligo; control hybridisation oligo; kinase labelling kit for oligo probe; and radioactive nucleotides. Additionally, bisulfite conversion reagents may include: DNA denaturation buffer; sulfonation buffer; DNA recovery reagents or kits (*e.g.*, precipitation, ultrafiltration, affinity column); desulfonation buffer; and DNA recovery components.

Typical reagents (*e.g.*, as might be found in a typical MethyLight®-based kit) for MethyLight® analysis may include, but are not limited to: PCR primers for specific gene (or methylation-altered DNA sequence or CpG island); TaqMan® probes; optimised PCR buffers and deoxynucleotides; and Taq polymerase.

Typical reagents (*e.g.*, as might be found in a typical Ms-SNuPE-based kit) for Ms-SNuPE analysis may include, but are not limited to: PCR primers for specific gene (or methylation-altered DNA sequence or CpG island); optimised PCR buffers and deoxynucleotides; gel extraction kit; positive control primers; Ms-SNuPE primers for specific gene; reaction buffer (for the Ms-SNuPE reaction); and radioactive nucleotides. Additionally, bisulfite conversion reagents may include: DNA denaturation buffer; sulfonation buffer; DNA recovery regents or kit (*e.g.*, precipitation, ultrafiltration, affinity column); desulfonation buffer; and DNA recovery components.

Typical reagents (*e.g.*, as might be found in a typical MSP-based kit) for MSP analysis may include, but are not limited to: methylated and unmethylated PCR primers for specific gene (or methylation-altered DNA sequence or CpG island), optimised PCR buffers and deoxynucleotides, and specific probes.

Specifically, the present invention is related to a method for characterising a cell proliferative disorder of the breast tissues and/or predicting the survival of a patient diagnosed with said disorder, comprising the steps of: a) detecting the expression of a nucleic acid or a polypeptide expressed from the PITX2 gene in an isolated biological sample representative of said cell proliferative disorders of the breast tissues and b) predicting therefrom the survival of said patient, characteristics of said cell proliferative disorder, and/or prognosis of said patient. Preferably, the method according to the present invention further comprises c) determining a suitable treatment regimen for the subject.

Preferred is a method according to the present invention, wherein the patient is characterised

by being subject to adjuvant endocrine therapy comprising one or more treatments which target the estrogen receptor pathway or are involved in estrogen metabolism, production or secretion.

Preferred is also a method according to the present invention, wherein said breast cell proliferative disorders are taken from the group comprising ductal carcinoma *in situ*, invasive ductal carcinoma, invasive lobular carcinoma, lobular carcinoma *in situ,* comedocarcinoma, inflammatory carcinoma, mucinous carcinoma, scirrhous carcinoma, colloid carcinoma, tubular carcinoma, medullary carcinoma, metaplastic carcinoma, and papillary carcinoma and papillary carcinoma *in situ*, undifferentiated or anaplastic carcinoma and Paget's disease of the breast.

According to another aspect of the method according to the present invention, said method is characterised in that the detection is carried out by a) amplification of a section of the gene PITX2 and/or microsatellites associated therewith; b) detecting the presence and/or absence of alleles of the amplificate; and c) predicting therefrom the survival of said patient, characteristics of said cell proliferative disorder, and/or prognosis of said patient.

According to another aspect of the method according to the present invention, said method is characterised in that the detection is carried out by a) contacting said biological sample with an antibody immunoreactive with the PITX2 polypeptide to form an immunocomplex; b) detecting said immunocomplex; and c) predicting therefrom the survival of said patient, characteristics of said cell proliferative disorder, and/or prognosis of said patient.

According to another aspect of the method according to the present invention, said method is characterised in that the detection is carried out by a) contacting said biological sample with an antibody immunoreactive with the PITX2 polypeptide to form an immunocomplex; b) detecting said immunocomplex; c) therefrom predicting the survival of said patient, characteristics of said cell proliferative disorder, and/or prognosis of said patient, and d) comparing the quantity of said immunocomplex to the quantity of immunocomplex formed under identical conditions with the same antibody and a control sample from one or more patients with a known prognosis.

According to yet another aspect of the method according to the present invention, the detection is carried out by a) contacting said biological sample with an antibody immunoreactive with the PITX2 polypeptide to form an immunocomplex; b) detecting said immunocomplex; and c) predicting therefrom the survival of said patient, characteristics of said cell proliferative disorder, and/or prognosis of said patient, and wherein an increase in quantity of said immunocomplex in the sample from said subject relative to said control sample is indicative of a bad prognosis.

Preferred is a method according to the present invention, wherein said immunoassay is a radioimmunoassay or an ELISA or a Western blot. Preferred is a method according to the present invention, wherein said detection is afforded by mRNA expression analysis. Most preferred is a method according to the present invention, comprising detecting the level of mRNA encoding a PITX2 polypeptide in a biological sample from a patient,

Preferred is a method according to the present invention, wherein a increased concentration of said mRNA above the concentration determined for an individual known to have a good prognosis indicates a bad prognosis.

According to yet another aspect of the method according to the present invention said method comprises the steps of: a) providing a polynucleotide probe which specifically hybridises or is identical to a polynucleotide consisting of SEQ ID NO: 19 or SEQ ID NO: 1, b) incubating said sample with said polynucleotide probe under high stringency conditions to form a specific hybridisation complex between a nucleic acid and said probe; and c) detecting said hybridisation complex.

Preferred is a method according to the present invention wherein said nucleic acid is mRNA or a cDNA derived therefrom.

Preferred is a method according to the present invention wherein the detecting step further comprises the steps of: a) producing a cDNA from mRNA in the sample; b) providing two oligonucleotides which specifically hybridise to regions flanking a segment of the cDNA; c) performing a polymerase chain reaction on the cDNA of step a) using the oligonucleotides of step b) as primers to amplify the cDNA segment; and d) detecting the amplified cDNA segment.

Yet another aspect of the present invention relates to a use of a polypeptide expressed from the PITX2 gene for differentiating or distinguishing between patients diagnosed with breast cancer, who have a good survival prognosis and patients who have a bad survival prognosis. Preferably, said polypeptide is expressed from the PITX2 gene and used for prediction of survival of a patient diagnosed with a cell proliferative disorder of the breast.

Preferred is a method according to the present invention wherein said detection comprises determining the genetic parameters of the gene PITX2, its promoter and/or regulatory elements. More preferred is a method according to the present invention wherein said detection comprises determining the epigenetic parameters of the gene PITX2, its promoter and/or regulatory elements.

Preferred is a method according to the present invention, wherein said detection comprises determining the methylation status of one or more CpG positions of a target nucleic acid within the gene PITX2, its promoter and/or regulatory elements, in particular through the methylation analysis of a genomic DNA sequence according to SEQ ID NO: 1. Preferred is further a method according to the present invention, wherein said detection comprises determining the methylation status of one or more CpG positions of a target nucleic acid characterised as being identical to or hybridising under stringent or moderately stringent conditions to a sequence out of the group of SEQ ID NO: NOs 13, 18 and 19. Preferred is further a method according to the present invention, wherein the methylation analysis is afforded by contacting said target nucleic acid with one or more agents that convert cytosine bases that are unmethylated at the 5'-position thereof to a base that is detectably dissimilar to cytosine in terms of hybridisation properties.

Preferred is a method according to the present invention, wherein contacting said target nucleic acids with one or more agents comprises use of a solution selected from the group consisting of bisulfite, hydrogen sulfite, disulfite, and combinations thereof.

Yet another aspect of the present invention relates to the use of a set of oligomer probes comprising at least two oligomers, in particular an oligonucleotide or peptide nucleic acid (PNA)-oligomer, said oligomer comprising in each case at least one base sequence having a length of at least 9 nucleotides which is complementary to, or hybridises under moderately stringent or

stringent conditions to a pretreated genomic DNA according to one of the SEQ ID NO: 2 to SEQ ID NO: 5 and sequences complementary thereto, for detecting the cytosine methylation state and/or single nucleotide polymorphisms (SNPs) within one of the sequences according to SEQ ID NO: 1, and sequences complementary thereto.

Yet another aspect of the present invention relates to a method for manufacturing an arrangement of different oligomers (array) fixed to a carrier material for analysing diseases associated with the methylation state of the CpG dinucleotides of one of SEQ ID NO: 1, and sequences complementary thereto wherein at least one oligomer, in particular an oligonucleotide or peptide nucleic acid (PNA)-oligomer, said oligomer comprising in each case at least one base sequence having a length of at least 9 nucleotides which is complementary to, or hybridises under moderately stringent or stringent conditions to a pretreated genomic DNA according to one of the SEQ ID NO: 2 to SEQ ID NO: 5 and sequences complementary thereto, is coupled to a solid phase.

Yet another aspect of the present invention relates to a composition of matter comprising the following: a) a nucleic acid comprising a sequence at least 18 bases in length of a segment of the chemically pretreated genomic DNA according to one of the sequences taken from the group comprising SEQ ID NO: 1 to SEQ ID NO: 5 and sequences complementary thereto, and b) a buffer comprising at least one of the following substances: 1 to 5 mM magnesium chloride, 100-500 µM dNTP, 0.5-5 units/10ul of taq polymerase, an oligomer, in particular an oligonucleotide or peptide nucleic acid (PNA)-oligomer, said oligomer comprising in each case at least one base sequence having a length of at least 9 nucleotides which is complementary to, or hybridises under moderately stringent or stringent conditions to a pretreated genomic DNA according to one of the SEQ ID NO: 2 to SEQ ID NO: 5 and sequences complementary thereto.

Preferably, the gene PITX2, its promoter and/or regulatory elements is used for predicting the survival of patients diagnosed with a cell proliferative disease. Preferred is the use of the mRNA of the gene PITX2 for predicting the survival of patients diagnosed with a cell proliferative disease.

Yet another aspect of the present invention relates to a method for predicting the survival of patients diagnosed with a cell proliferative disease according to the present invention, com-

prising: a) isolating or enriching genomic DNA from said biological sample ; b) treating the genomic DNA, or a fragment thereof, with one or more reagents to convert 5-position unmethylated cytosine bases to uracil or to another base that is detectably dissimilar to cytosine in terms of hybridisation properties; c) contacting the treated genomic DNA, or the treated fragment thereof, with an amplification enzyme and at least two primers comprising, in each case a contiguous sequence at least 18 nucleotides in length that is complementary to, or hybridises under moderately stringent or stringent conditions to a sequence selected from the group consisting of SEQ ID NO: 2 to 5, and complements thereof, wherein the treated DNA or a fragment thereof is either amplified to produce one or more amplificates, or is not amplified; and d) determining, based on the presence or absence of, or on the quantity or on a property of said amplificate, the methylation state of at least one CpG dinucleotide sequence of SEQ ID NO: 1, or an average, or a value reflecting an average methylation state of a plurality of CpG dinucleotide sequences of SEQ ID NO: 1.

Yet another aspect of the present invention relates to a method for detecting the survival of patients diagnosed with a cell proliferative disease of the breast according to the present invention, comprising the following steps of a) obtaining, from a subject, a biological sample having subject genomic DNA; b) treating the genomic DNA, or a fragment thereof, with one or more reagents to convert 5-position unmethylated cytosine bases to uracil or to another base that is detectably dissimilar to cytosine in terms of hybridisation properties; c) amplifying one or more fragments of the treated DNA such that only DNA originating from breast or breast cell proliferative disorder cells are amplified and d) detecting the amplificates or characteristics thereof and thereby deducing on the survival of patients diagnosed with a cell proliferative disease of the breast.

Yet another aspect of the present invention relates to a use of an oligomer, an oligonucleotide or peptide nucleic acid (PNA)-oligomer, said oligomer comprising in each case at least one base sequence having a length of at least 9 nucleotides which is complementary to, or hybridises under moderately stringent or stringent conditions to an artificially modified, chemically pretreated DNA according to one of the SEQ ID NO: 2 to SEQ ID NO: 5 and sequences complementary thereto, for differentiating or distinguishing between patients diagnosed with breast cancer, who have a good survival prognosis and patients who have a bad survival prognosis.

Yet another aspect of the present invention relates to a use of a nucleic acid comprising a sequence of at least 18 bases in length of a segment of the artificially modified, chemically pretreated, DNA according to one of the sequences taken from the group comprising SEQ ID NO: 2 to SEQ ID NO: 5 and sequences complementary thereto, for differentiating or distinguishing between patients diagnosed with breast cancer, who have a good survival prognosis and patients who have a bad survival prognosis. Yet another aspect of the present invention relates to a use of an oligomer, an oligonucleotide or peptide nucleic acid (PNA)-oligomer, said oligomer comprising in each case at least one base sequence having a length of at least 9 nucleotides which is complementary to, or hybridises under moderately stringent or stringent conditions to an artificially modified, chemically pretreated DNA according to one of the SEQ ID NO: 2 to SEQ ID NO: 5 and sequences complementary thereto, for prediction of survival of a patient diagnosed with a cell proliferative disorder of the breast.

Yet another aspect of the present invention relates to a use of a nucleic acid comprising a sequence of at least 18 bases in length of a segment of the artificially modified, chemically pretreated, DNA according to one of the sequences taken from the group comprising SEQ ID NO: 2 to SEQ ID NO: 5 and sequences complementary thereto, for prediction of survival of a patient diagnosed with a cell proliferative disorder of the breast. Yet another aspect of the present invention relates to a use of a nucleic acid represented by their SEQ ID NO: NO out of the group of nucleic acids according to SEQ ID NO: NOS : 6, 7, 8, 9, 10, 14, 15, 22, 23, 24, 25, 26, 27, 28, 29, 30 and 31, for differentiating or distinguishing between patients diagnosed with breast cancer, who have a good survival prognosis and patients who have a bad survival prognosis. Yet another aspect of the present invention relates to a use of a nucleic acid represented by their SEQ ID NO: NO out of the group of nucleic acids according to SEQ ID NO: NOS : 6, 7, 8, 9, 10, 14, 15, 22, 23, 24, 25, 26, 27, 28, 29, 30 and 31, for prediction of survival of a patient diagnosed with a cell proliferative disorder of the breast.

In the context of this invention the terms "obtaining a biological sample" or "obtaining a sample from a subject", is supposed to comprise several different sources of such a sample, but always excludes the active retrieval of a sample from an individual patient, such as the performance of a biopsy. Included are the following examples: obtaining a sample, which was prior to the obtaining step taken from a patient in a biopsy or surgery, from a sample provider; obtaining a sample from a clinician, a surgeon or other medical personnel; obtaining a sample from a courier, who is bringing the sample from the clinician or practitioner or patient him-

self, for example to the analytic service station, as well as obtaining a sample such as a body fluid sample per post or from the hands of the patient himself. This is not meant to be a limiting list, but shall illustrate that it is not a feature of the invention that it needs to be carried out on the patient itself.

The term "biological material" relates to any material that is derived from a source, in particular an animal and/or human source, that contains or is suspected to contain genomic DNA. On example of a biological material according to the present invention will be a biological sample.

In the context of the present invention, the term "CpG island" refers to a contiguous region of genomic DNA that satisfies the criteria of (1) having a frequency of CpG dinucleotides corresponding to an "Observed/Expected Ratio" >0.6, and (2) having a "GC Content" >0.5. CpG islands are typically, but not always, between about 0.2 to about 1 kb in length.

In the context of the present invention the term "regulatory region" of a gene is taken to mean nucleotide sequences which affect the expression of a gene. Said regulatory regions may be located within, proximal or distal to said gene. Said regulatory regions include but are not limited to constitutive promoters, tissue-specific promoters, developmental-specific promoters, inducible promoters and the like. Promoter regulatory elements may also include certain enhancer sequence elements that control transcriptional or translational efficiency of the gene.

In the context of the present invention, the term "methylation" refers to the presence or absence of 5-methylcytosine ("5-mCyt") at one or a plurality of CpG dinucleotides within a DNA sequence.

In the context of the present invention the term "methylation state" is taken to mean the degree of methylation present in a nucleic acid of interest, this may be expressed in absolute or relative terms i.e. as a percentage or other numerical value or by comparison to another tissue and therein described as hypermethylated, hypomethylated or as having significantly similar or identical methylation status.

In the context of the present invention, the term "hemi-methylation" or "hemimethylation" refers to the methylation state of a palindromic CpG methylation site, where only a single

cytosine in one of the two CpG dinucleotide sequences of the double stranded CpG methylation site is methylated (*e.g.*, 5'-NNC^{M}GNN-3' (top strand): 3'-NNGCNN-5' (bottom strand)).

In the context of the present invention, the term "hypermethylation" refers to the average methylation state corresponding to an *increased* presence of 5-mCyt at one or a plurality of CpG dinucleotides within a DNA sequence of a test DNA sample, relative to the amount of 5-mCyt found at corresponding CpG dinucleotides within a normal control DNA sample.

In the context of the present invention, the term "hypomethylation" refers to the average methylation state corresponding to a *decreased* presence of 5-mCyt at one or a plurality of CpG dinucleotides within a DNA sequence of a test DNA sample, relative to the amount of 5-mCyt found at corresponding CpG dinucleotides within a normal control DNA sample.

In the context of the present invention, the term "microarray" refers broadly to both "DNA microarrays," and 'DNA chip(s),' as recognised in the art, encompasses all art-recognised solid supports, and encompasses all methods for affixing nucleic acid molecules thereto or synthesis of nucleic acids thereon.

"Genetic parameters" are mutations and polymorphisms of genes and sequences further required for their regulation. To be designated as mutations are, in particular, insertions, deletions, point mutations, inversions and polymorphisms and, particularly preferred, SNPs (single nucleotide polymorphisms).

"Epigenetic modifications" or "epigenetic parameters" are modifications of DNA bases of genomic DNA and sequences further required for their regulation, in particular, cytosine methylations thereof. Further epigenetic parameters include, for example, the acetylation of histones which, however, cannot be directly analysed using the described method but which, in turn, correlate with the DNA methylation.

In the context of the present invention, the term "bisulfite reagent" refers to a reagent comprising bisulfite, disulfite, hydrogen sulfite or combinations thereof, useful as disclosed herein to distinguish between methylated and unmethylated CpG dinucleotide sequences.

In the context of the present invention, the term "Methylation assay" refers to any assay for determining the methylation state of one or more CpG dinucleotide sequences within a sequence of DNA.

In the context of the present invention, the term "MS.AP-PCR" (Methylation-Sensitive Arbitrarily-Primed Polymerase Chain Reaction) refers to the art-recognised technology that allows for a global scan of the genome using CG-rich primers to focus on the regions most likely to contain CpG dinucleotides, and described by Gonzalgo et al., *Cancer Research* 57:594-599, 1997.

In the context of the present invention, the term "MethyLight®" refers to the art-recognised fluorescence-based real-time PCR technique described by Eads et al., *Cancer Res.* 59:2302-2306, 1999.

In the context of the present invention, the term "HeavyMethyl™" assay, in the embodiment thereof implemented herein, refers to a HeavyMethyl™ MethylLight assay, which is a variation of the MethylLight assay, wherein the MethylLight assay is combined with methylation specific *blocking* probes covering CpG positions between the amplification primers.

The term "Ms-SNuPE" (Methylation-sensitive Single Nucleotide Primer Extension) refers to the art-recognised assay described by Gonzalgo & Jones, *Nucleic Acids Res.* 25:2529-2531, 1997.

In the context of the present invention the term "MSP" (Methylation-specific PCR) refers to the art-recognised methylation assay described by Herman et al. *Proc. Natl. Acad. Sci. USA* 93:9821-9826, 1996, and by US Patent No. 5,786,146.

In the context of the present invention the term "COBRA" (Combined Bisulfite Restriction Analysis) refers to the art-recognised methylation assay described by Xiong & Laird, *Nucleic Acids Res.* 25:2532-2534, 1997.

In the context of the present invention the term "hybridisation" is to be understood as a bond of an oligonucleotide to a complementary sequence along the lines of the Watson-Crick base pairings in the sample DNA, forming a duplex structure.

"Stringent hybridisation conditions," as defined herein, involve hybridising at 68°C in 5x SSC/5x Denhardt's solution/1.0% SDS, and washing in 0.2x SSC/0.1% SDS at room temperature, or involve the art-recognised equivalent thereof (e.g., conditions in which a hybridisation is carried out at 60°C in 2.5 x SSC buffer, followed by several washing steps at 37°C in a low buffer concentration, and remains stable). Moderately stringent conditions, as defined herein, involve including washing in 3x SSC at 42°C, or the art-recognised equivalent thereof. The parameters of salt concentration and temperature can be varied to achieve the optimal level of identity between the probe and the target nucleic acid. Guidance regarding such conditions is available in the art, for example, by Sambrook et al., 1989, Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, N.Y.; and Ausubel et al. (eds.), 1995, Current Protocols in Molecular Biology, (John Wiley & Sons, N.Y.) at Unit 2.10.

"Background DNA" as used herein refers to any nucleic acids which originate from sources other than colon cells.

In the context of this application "survival" is meant to describe the time from diagnosis or start of treatment to an endpoint, which may be either the time of death (considering any reason for death or only death from breast cancer), or the time of recurrence of breast cancer (for example in form of metastases), which may be local or distant, or the time of occurrence of any breast cancer associated disease. Therefore "predicting the survival" is meant to comprise predicting the disease free survival, as well as the overall survival or any other consideration of time between diagnosis and endpoint of treatment. However, as it is obvious in the state of the art, a precise prediction of life time is generally impossible whether it is based on a biomarker analysis, or any other prognostic tools, it is understood throughout the invention that said term "prediction of survival" or "predicting the survival" is used to describe the risk of patient to suffer from a recurrence of metastasis or other disease caused by the original breast cell proliferative disease the patient was diagnosed with (also termed "risk of relapse"). Said risk can be predicted with a certain probability or likelihood. It is also clear, that predicting the survival is meant to comprise the determination of the likelihood or probability whether a subject or patient will survive for a longer or shorter period of time.

Throughout this invention it is preferred that said survival is characterised as the disease free or the overall survival. It is especially preferred that survival is understood as disease free

survival. Disease free survival is understood as absence of recurrence of cancer (local or distant).

The terms "endocrine therapy" or "endocrine treatment" is meant to comprise any therapy, treatment or treatments targeting the estrogen receptor pathway or estrogen synthesis pathway or estrogen conversion pathway, which is involved in estrogen metabolism, production or secretion. Said treatments include, but are not limited to estrogen receptor modulators, estrogen receptor down-regulators, aromatase inhibitors, ovarian ablation, LHRH analogues and other centrally acting drugs influencing estrogen production.

The term "monotherapy" is used to explain that no other treatment is given in addition or to support said monotherapy.

In the context of the present invention the term "chemotherapy" is taken to mean the use of drugs or chemical substances to treat cancer. This definition excludes radiation therapy (treatment with high energy rays or particles), hormone therapy (treatment with hormones or hormone analogues (synthetic substitutes) and surgical treatment.

In the context of the present invention the term "adjuvant treatment" is taken to mean a therapy of a cancer patient immediately following an initial non chemotherapeutical therapy, e.g. surgery. In general, the purpose of an adjuvant therapy is to provide a significantly smaller risk of recurrences compared without the adjuvant therapy.

In the context of the present invention the term "determining a suitable treatment regimen for the subject" is taken to mean a treatment regimen (i.e. a single therapy or a combination of different therapies that are used for the prevention and/or treatment of the cancer in the patient) for the cancer patient that is started, modified and/or ended based or essentially based or at least partially based on the results of the analysis according to the present invention. One example is starting an adjuvant endocrine therapy after surgery, another would be to modify the dosage of a particular chemotherapy. The determination can, in addition to the results of the analysis according to the present invention, be based on personal characteristics of the subject to be treated. In most cases, the actual determination of the suitable treatment regimen for the subject will be performed by the attending physician or doctor.

While the present invention has been described with specificity in accordance with certain of its preferred embodiments, the following examples and figures serve only to illustrate the invention and is not intended to limit the invention within the principles and scope of the broadest interpretations and equivalent configurations thereof.

In the accompanying sequence protocol and the figures,

SEQ ID NO: 1 shows the nucleic acid sequence of the human gene PITX2,

SEQ ID NO: 2 to 5 show chemically pretreated nucleic acid sequences of the gene PITX2, according to table 1.

SEQ ID NO: 6 to 9 show the nucleic acid sequences of those primers and probes useful to predict the survival of breast cancer patients according to the invention as described in example 4.

SEQ ID NO: 14 to 17 show the nucleic acid sequences of those primers and probes useful to predict the survival of breast cancer patients according to the invention as described in example 5.

SEQ ID NO: 10 to 12 show the nucleic acid sequences of primers and probes according to a control gene used in the example 4 and 5.

SEQ ID NO: 13 shows a subsequence of SEQ ID NO: 1, which represents the nucleic acid sequence of the human gene PITX2.

SEQ ID NO: 18 shows an amino acid sequence of the polypeptide encoded by the gene PITX2. The amino acid sequence of the polypeptide encoded by the gene PITX2 is also illustrated in figure 10.

FIGURES

Figure 1 presents a scheme to illustrate a preferred application of the method according to the invention. Along the Y axis tumour(s) mass (or size) increases, wherein the line '3' indicates the limit of detectability of said tumour mass. The X axis represents time (such as in life time of a patient). Accordingly said figure illustrates a simplified model of an Stage 1-3 breast tumour wherein primary treatment was surgery (at point 1), followed by adjuvant therapy with Tamoxifen, as an example for an endocrine treatment. In a first scenario a patient without relapse during endocrine treatment (4) is shown as remaining below the limit of detectability for the duration of the observation. A patient with relapse of the cancer (5) has a period of disease free survival (2) followed by relapse when the carcinoma mass reaches the level of detectability.

Figure 2 shows the result of the assay (QM assay) as described in example 4: A Kaplan-Meier estimated metastasis-free survival curve for 3 CpG sites of the PITX2 gene by means of Real-Time methylation specific probe analysis (QM assay). The lower curve shows the proportion of metastasis free patients in the population with above median methylation levels, the upper curve shows the proportion of metastasis free patients in the population with below median methylation levels. The X axis shows the metastasis free survival times of the patients in months, and the Y axis shows the proportion of metastasis free survival patients.

Figure 3 shows the result of the chip hybridisation experiment as described in example 2. A Kaplan-Meier estimated metastasis-free survival curves for 2 CpG positions of the PITX2 gene by means of methylation specific detection oligo hybridisation analysis. The lower curve shows the proportion of metastasis free patients in the population with above median methylation levels, the upper curve shows the proportion of metastasis free patients in the population with below median methylation levels. The X axis shows the metastasis free survival times of the patients in months, and the Y axis shows the proportion of metastasis free survival patients.

Figure 4 shows the Kaplan-Meier estimated metastasis-free survival curves for 2 CpG positions of the PITX2 gene by means of methylation specific detection oligo hybridisation analysis. The lower line shows the proportion of metastasis free patients in the population of 55 patients with above median methylation levels, the upper curve shows the proportion of metastasis free patients in the population of 54 patients with below median methylation levels. The X axis shows the metastasis free survival times of the patients in years, and the Y axis shows the proportion of metastasis free survival patients in %. This resulted from a first data set that was achieved in a first study.

Figure 5 shows the Kaplan-Meier estimated metastasis-free survival curves for 6 different CpG positions located within the preferred region of the PITX2 gene (SEQ ID NO: 13) by means of methylation specific detection oligo hybridisation analysis. The lower line shows the proportion of metastasis free patients in the population of 118 patients with above median methylation levels, the upper curve shows the proportion of metastasis free patients in the population of 118 patients with below median methylation levels. The X axis shows the metastasis free survival times of the patients in years, and the Y axis shows the proportion of

metastasis free survival patients in %. This resulted from a second data set that was achieved in a second study.

Figure 6 shows the Kaplan-Meier estimated metastasis-free survival curves for 6 different CpG positions located within the preferred region of the PITX2 gene (SEQ ID NO: 13) by means of methylation specific detection oligo hybridisation analysis. This time only a sub-population of 148 patients, characterised by a tumour at grade G1 or G2, was analysed: The lower curve shows the proportion of metastasis free patients in the population of 74 patients with above median methylation levels, the upper curve shows the proportion of metastasis free patients in the population of 74 patients with below median methylation levels. The X axis shows the metastasis free survival times of the patients in years, and the Y axis shows the proportion of metastasis free survival patients in %. This resulted from a second data set that was achieved in the second example.

Figure 7 shows the Kaplan-Meier estimated metastasis-free survival curves for 4 different CpG positions located within the preferred region of the PITX2 gene (SEQ ID NO: 13) by means of methylation specific detection oligo hybridisation analysis. This time a subpopulation of 224 patients, characterised by a tumour of stage 1 or 2 (T1 or T2), was analysed: The lower curve shows the proportion of metastasis free patients in the population of 112 patients with above median methylation levels, the upper curve shows the proportion of metastasis free patients in the population of 112 patients with below median methylation levels. The X axis shows the metastasis free survival times of the patients in years, and the Y axis shows the proportion of metastasis free survival patients in %. This resulted from the second data set that was achieved in the second example.

Figure 8 shows the disease-free survival curves for a combination of two oligonucleotides each from the genes TBC1D3 and CDK6, and one oligonucleotide from the gene PITX2 covering two CpG sites. The black curve shows the proportion of disease free patients in the population with above median methylation scores, the grey curve shows shows the proportion of disease free patients in the population with below median methylation scores.

Figure 9 shows the plot according to Figure 8 and the classification of the sample set by means of the St. Gallen method. The unbroken lines represent the methylation analysis wherein the black curve shows the proportion of disease free patients in the population with

above median methylation scores, the grey curve shows the proportion of disease free patients in the population with below median methylation scores. The broken lines represent the St. Gallen classification of the sample set wherein the black curve shows the disease free survival time of the high risk group and the grey curve shows the disease free survival of the low risk group.

Figure 10 illustrates the amino acid sequence of the polypeptide encoded by the gene PITX2.

EXAMPLES

EXAMPLE 1 : Study 1

The first study was based on a population of 109 patients, comprising patients of both nodal statuses N0 and N+. All patients were ER+ (estrogen receptor positive). All patients received Tamoxifen monotherapy immediately after surgery or diagnosis. The samples were analysed using Epigenomics' chip technology with two chip panels representing altogether 117 candidate genes. For further details see patent application WO 04/035803 and EP 03 090 432.0, which is hereby incorporated by reference. In this study one significant marker gene was found. The methylation status of PITX2, coding for a transcription factor, was correlated statistically significant with disease-free survival under adjuvant Tamoxifen treatment. A Cox regression model that includes the nodal status of the patient at the time of diagnosis was applied.

The result from this study – with respect to PITX2 - is illustrated in Figure 4. The X axis shows the metastasis free survival times of the patients in years, and the Y axis shows the proportion of metastasis free survival patients in %. Amongst the 54 patients (upper line) with below median methylation levels a higher percentage has a significantly longer metastasis free survival time, than amongst the 55 patients with above median methylation levels (lower line). To illustrate the result: At time of 10 years after surgery under tamoxifen monotherapy, more than 75% of the patients with low methylation in PITX2 are still metastasis free, whereas less than 60% of the patients with high methylation in PITX2.

As the survival of a breast cancer patient is known to also be correlated to the patient's nodal status, the differentiating power of the marker in this mixed population is expected to be less than in a homogenous population.

Another study was performed to analyse whether the same marker can be identified independently, in a completely different set of patient samples and also to characterise the differential power towards predicting survival for a sub-group of patients, all being N0.

EXAMPLE 2 : Study 2

The second study was based on samples from 236 patients from 5 different centres, wherein all patients were N0 (nodal status negative), and older than 35 years. In all cases the surgery was performed before 1998. All patients were ER+ (estrogen receptor positive), and the tumours were graded to be T1-3, G1-3. In this study all patients received Tamoxifen directly after surgery, and the outcome was assessed as the length of disease-free survival. In order to be as representative as possible for the final target group, the patients and their tumour samples had to fulfil the following criteria:

The range and median follow-up of patients were the following:
Median: 64.5 months
Range: 3 months to 142 months
(calculated based on patients who were disease-free at end of observation time).

Analysis of the methylation patterns of patient samples treated with Tamoxifen as an adjuvant therapy immediately following surgery (see Figure 1) is shown in the plots according to Figures 5 to 7. For the amplificate, the mean methylation over 4 oligo-pairs for that amplificate was calculated and the population split into groups according to their mean methylation values, wherein one group was composed of individuals with a methylation score higher than the median and a second group composed of individuals with a methylation score lower than the median.

The primer oligonucleotides used to generate the amplificate, that was analysed in the array experiment were these :
Array Primer PITX2_Q21: GTAGGGGAGGGAAGTAGATGT (SEQ ID NO: 22)
Array Primer PITX2_R23: TCCTCAACTCTACAAACCTAAAA (SEQ ID NO: 23)
The according genomic region of said amplificate is given in SEQ ID NO: 13.

The sequences of the oligonucleotides used in this array experiment were the following:

SEQ ID NO: NO 24 : AGTCGGGAGAGCGAAA

SEQ ID NO: NO 25 : AGTTGGGAGAGTGAAA

SEQ ID NO: NO 26 : AAGAGTCGGGAGTCGGA

SEQ ID NO: NO 27:_AAGAGTTGGGAGTTGGA

SEQ ID NO: NO 28: GGTCGAAGAGTCGGGA

SEQ ID NO: NO 29: GGTTGAAGAGTTGGGA

SEQ ID NO: NO 30: ATGTTAGCGGGTCGAA

SEQ ID NO: NO 31: TAGTGGGTTGAAGAGT

When the data derived from analysing 6 different CpG sites, located within the preferred amplified region of the PITX2 gene by means of methylation specific detection oligo hybridisation analysis were plotted as Kaplan-Meier estimated metastasis-free survival curves, it can be seen that the differential power of the marker PITX2 increased with selecting for N0 patients. This is shown in figures 5 to 7. The X axis shows the metastasis free survival times of the patients in years, and the Y axis shows the proportion of metastasis free survival patients in %. The lower curve shows the proportion of metastasis free patients in the population with above median methylation levels, and the upper curve shows the proportion of metastasis free patients in the population with below median methylation levels.

For example, as illustrated in figure 5, 10 years after surgery only about 65% of the patients of the 118 patients with the higher methylation status are metastasis free, whereas about 90% of the 118 patients with lower methylation status are metastasis free.

As illustrated in figure 6 when looking at the analogous Kaplan-Meier analysis for a subpopulation of 148 patients, characterised by a tumour at stage G1 or G2 this differential power increases again: 10 years after surgery only about 60% of the 74 patients with the higher methylation status are metastasis free, whereas about 95% of the 74 patients with lower methylation status are metastasis free.

Figure 7 illustrates how the survival is also correlated to the tumour stage at surgery by showing the analogous Kaplan-Meier analysis for a subpopulation of 150 patients, characterised by a tumour stage of T1 or T2: The number of patients with 10 years MFS is about 68% of patients of the 112 with the higher methylation status, whereas about 95% of the 112 patients with lower methylation status are metastasis free.

EXAMPLE 3:

The accuracy of the differentiation between the different groups was further increased by combining multiple oligonucleotides from different genes. As described in the text it was recognised that adding additional informative markers to the analysis could potentially increase the prognostic power of a survival test. Therefore it was calculated how a combination of two methylation specific oligonucleotides each from the genes TBC1D3 and CDK6, and one oligonucleotide from the gene PITX2 would differentiate the groups of good or bad prognosis. The result is shown in figure 8 as the according Kaplan-Meier curve.

Figure 9 shows -on top of Figure 8- the classification of the patients from the sample set by means of the St.Gallen method (the current method of choice for estimating disease free survival), thereby showing the improved effectiveness of methylation analysis over current methods, in particular post 80 months.

EXAMPLE 4: Real time Quantitative methylation analysis

Genomic DNA was analysed using the Real Time PCR technique after bisulfite conversion. In this analysis four oligonucleotides were used in each reaction. Two non methylation specific PCR primers were used to amplify a segment of the treated genomic DNA containing a methylation variable oligonucleotide probe binding site. Two oligonucleotide probes competitively hybridise to the binding site, one specific for the methylated version of the binding site, the other specific to the unmethlyated version of the binding site. Accordingly, one of the probes comprises a CpG at the methylation variable position (i.e. anneals to methylated bisulphite treated sites) and the other comprises a TpG at said position (i.e. anneals to unmethylated bisulphite treated sites). Each species of probe is labelled with a 5' fluorescent reporter dye and a 3' quencher dye wherein the CpG and TpG oligonucleotides are labelled with different dyes.

The reactions are calibrated by reference to DNA standards of known methylation levels in order to quantify the levels of methylation within the sample. The DNA standards were composed of bisulfite treated phi29 amplified genomic DNA (i.e. unmethlyated), and/or phi29 amplified genomic DNA treated with Sss1 methylase enzyme (thereby methylating each CpG position in the sample), which is then treated with bisulfite solution. Seven different reference

standards were used with 0%, (i.e. phi29 amplified genomic DNA only), 5%, 10%, 25%, 50%, 75% and 100% (i.e. phi29 Sss1 treated genomic only).

The amount of sample DNA amplified is quantified by reference to the gene (ß-actin (ACTB)) to normalise for input DNA. For standardisation the primers and the probe for analysis of the ACTB gene lack CpG dinucleotides so that amplification is possible regardless of methylation levels. As there are no methylation variable positions, only one probe oligonucleotide is required.

The following oligonucleotides were used in the reaction to amplify the control amplificate:
Control Primer1: TGGTGATGGAGGAGGTTTAGTAAGT (SEQ ID NO: 10)
Control Primer2: AACCAATAAAACCTACTCCTCCCTTAA (SEQ ID NO: 11)
Control Probe: 6FAM-ACCACCACCCAACACACAATAACAAACACA-TAMRA or Dabcyl (SEQ ID NO: 12)

The nucleic acid sequence of the gene PITX2 is given in (SEQ ID NO: 1), after treatment with bisulfite two different strands are generated, and each of the strands is represented twice, once in a prior to treatment methylated version (SEQ ID NO: 2 and 3) and once in the prior to treatment unmethylated form (SEQ ID NO: 4 and 5), which are characterised as containing no cytosine bases (despite of those 5' adjacent to a guanine and methylated before treatment).
The following primers are used to generate an amplificate within the PITX2 sequence comprising the CpG sites of interest:

Primers for PITX bisulfite amplificate length : 144 bp
PITX2R02: GTAGGGGAGGGAAGTAGATGTT (SEQ ID NO: 6)
PITX2Q02: TTCTAATCCTCCTTTCCACAATAA (SEQ ID NO: 7)

The genomic region according to the generated amplifacte of 144 bp in length is given in SEQ ID NO: NO 18.
Probes:
PITX2cg1: FAM-AGTCGGAGTCGGGAGAGCGA-Darquencher (SEQ ID NO: 8)
As an alternative quencher TAMRA was also used in additional experiments:
FAM-AGTCGGAGTCGGGAGAGCGA-TAMRA

PITX2tg1: YAKIMA YELLOW-AGTTGGAGTTGGGAGAGTGAAAGGAGA-Darquencher (SEQ ID NO: 9)

In additional experiments we also used :

VIC- AGTTGGAGTTGGGAGAGTGAAAGGAGA -TAMRA

The extent of methylation at a specific locus was determined by the following formula:

methylation rate = 100 * I (CG) / (I(CG) + I(TG))

(I = Intensity of the fluorescence of CG-probe or TG-probe)

PCR components were ordered from Eurogentec :

3 mM MgCl2 buffer, 10x buffer, Hotstart TAQ

Program (45 cycles): 95 °C, 10 min; 95 °C, 15 sec; 62 °C, 1 min

This assay was performed on 236 samples identical to those used in Example 2. The result is shown in figure 2. Figure 2 shows the Kaplan-Meier estimated disease-free survival curves for 3 CpG positions of the PITX2 gene by means of Real-Time (RT) methylation specific probe analysis, as described above. The lower curve shows the proportion of disease free patients in the population with above median methylation levels, the upper curve shows the proportion of disease free patients in the population with below median methylation levels. The X axis shows the disease free survival times of the patients in months, and the Y- axis shows the proportion of disease free survival patients. The p-value (probability that the observed distribution occurred by chance) was calculated as 0.0031, thereby confirming the data obtained by means of array analysis.

For comparison, figure 3 illustrates the result from the array analysis of said gene, according to the chip hybridisation experiment described in Example 2, wherein detection oligos were used (for details see EP 03 090 432.0, which is incorporated by reference). The p-value (probability that the observed distribution occurred by chance) was calculated as 0.0011.

EXAMPLE 5

Another QM assay was developed in our hands, which also performed very well. The following PITX2 specific oligonucleotides were employed to generate an amplificate of 164 bp. The oligonucleotides are specific for three co-methylated CpG positions:

Primers for PITX2 bisulfite amplificate with a length of 162 bp :

PITX2O2: AACATCTACTTCCCTCCCCTAC (SEQ ID NO: 14)

PITX2P3: GTTAGTAGAGATTTTATTAAATTTTATTGTAT (SEQ ID NO: 15)

The genomic region according to the generated amplificate of 162 bp in length is given in SEQ ID NO: NO 19.

Probes (from ABI):

PITX2-IIcg1: FAM-TTCGGTTGCGCGGT-MGBNQF (SEQ ID NO: 16)

PITX2-IItg1: VIC-TTTGGTTGTGTGGTTG- MGBNQF (SEQ ID NO: 17)

The extent of methylation at a specific locus was determined by the following formula:

methylation rate = 100 * I (CG) / (I(CG) + I(TG))

(I = Intensity of the fluorescence of CG-probe or TG-probe)

PCR components were ordered from Eurogentec : 2,5 mM MgCl2 buffer, 10x buffer, Hotstart TAQ

Program (45 cycles): 95 °C, 10 min; 95 °C, 15 sec; 60 °C, 1 min

Example 6

Patient material

The material to be used in this study, consists of fresh frozen healthy breast tissue, fresh frozen breast tumour tissue from untreated breast cancer patients (follow up over >10 years) and samples from Tamoxifen treated patients (follow up over >10 years from Tamoxifen treatment). Aliquots of DNA from these microdissected lesions are used as the source template for PCR-based LOH (Loss of heterozygosity) analysis. All tumour samples were derived from ER+ node negative patients.

LOH analysis

DNA from all tissue samples is subjected to PCR-based LOH analysis using two 4q25-26 markers (D4S1284 and D4S406). These markers define a region on chromosome 4 comprising the gene PITX2 gene said region but being more than 8.5 kbp distant of a region previously shown to undergo LOH in breast carcinomas [*Cancer Research* 59, 3576-3580, August 1, 1999].

DNA Extraction

Extract DNA from samples using the Wizzard Kit (Promega).

PCR reaction

See Clin. Cancer Res., *5:* 17-23, 1999 for further details.

Analyse each sample by means of single-plex PCR using the following primers:

D4S406

Forward primer: GAAAGGCAGAGTCATAACAGGAAG (SEQ ID NO: 32)

Reverse primer: TAAGGATAGAGTGATTTCCAAGAAAG (SEQ ID NO: 33)

PCR product size:205 (bp)

GenBank Accession:Z16728

D4S1284

Forward primer: CTTATCTGACAACAAGCGAGTATG (SEQ ID NO: 34)

Reverse primer: CAATTATTGTATTGTAGCATCGGAG (SEQ ID NO: 35)

PCR product size:172 (bp)

GenBank Accession:L14168

Synthesise forward primers with either a fluorescent FAM tag (D4S1284) or a fluorescent TET tag (D4S406) at the 5' end.

Prepare a suitable quantity of nucleotide mixture according to Table 2.

Aliquot 1 μl of each DNA sample into separate PCR tubes, add 9 μl reaction mixture according to Table 3 and thermal cycle according to the following conditions.

Thermal cycling conditions:

95°C for 15 min

39 cycles:

95°C for 1 min

55°C for 0:45

72°C for 1:15

72°C for 10 min

Gel electrophoresis

Horizontal ultrathin, high throughput fluorescence-based DNA fragment gel electrophoresis is the preferred technique to separate and analyse the PCR-generated alleles. Combine one microliter of amplified material with 2 µl formamide loading dye (APB) prior to electrophoresis. Add ROX 350 fluorescent size markers (0.7 µl; ABI) to amplified tumour DNA to allow sizing of alleles.

Heat samples to 95°C, load on 70 µm, 5% horizontal polyacrylamide gel and electrophorese for 1 h and 15 min at 30 W in 1 × TBE.

Data may be collected as commonly known in the art (see for example Clin. Cancer Res., *5*: 17-23, 1999). To determine whether allelic deletion had occurred at individual markers, calculate allelic ratios and express as a percentage of loss of intensity for the treated and untreated tumour samples compared with the corresponding normal samples (D-value) after normalisation. When the allelic ratio in the tumour DNA is reduced by greater than 40% (DO.40) from that found in the normal DNA, the sample is denoted as having LOH at that locus.

Table 2: Nucleotide Mix

10 µl dATP, 10 mM
10 µl dGTP, 10 mM
10 µl dTTP, 10 mM
2.0 µl dCTP, 10 mM
288 µl DEPC-treated H_2O

Table 3: Reaction mixture

1.0 µl	Taq Buffer
0.8 µl	Reduced nucleotide mixture
0.2 µl	Forward primer, 20 µM
0.2 µl	Reverse primer, 20 µM
6.6 µl	DEPC treated H_20
0.1 µl	γ-32P dCTP
0.1 µl	AmpliTaq Gold Polymerase

Total volume = 9 µl

CLAIMS

1. A method for characterising a cell proliferative disorder of the breast tissues and/or predicting the survival of a patient diagnosed with said disorder, comprising the steps of:

(a) detecting the expression of a nucleic acid or a polypeptide expressed from the PITX2 gene in an isolated biological sample representative of said cell proliferative disorders of the breast tissues and

(b) therefrom predicting the survival of said patient, characteristics of said cell proliferative disorder, and/or prognosis of said patient.

2. The method according to claim 1 further comprising

(c) determining a suitable treatment regimen for the subject.

3. The method of claim 1, wherein said patient is characterised by being subject to adjuvant endocrine therapy comprising one or more treatments which target the estrogen receptor pathway or are involved in estrogen metabolism, production or secretion.

4. The method of claim 1, wherein said breast cell proliferative disorders are taken from the group comprising ductal carcinoma *in situ*, invasive ductal carcinoma, invasive lobular carcinoma, lobular carcinoma *in situ,* comedocarcinoma, inflammatory carcinoma, mucinous carcinoma, scirrhous carcinoma, colloid carcinoma, tubular carcinoma, medullary carcinoma, metaplastic carcinoma, and papillary carcinoma and papillary carcinoma *in situ*, undifferentiated or anaplastic carcinoma and Paget's disease of the breast.

5. The method according to claim 1 characterised in that the detection is carried out by

a) contacting said biological sample with an antibody immunoreactive with the PITX2 polypeptide to form an immunocomplex;

b) detecting said immunocomplex; and

c) predicting therefrom the survival of said patient, characteristics of said cell proliferative disorder, and/or prognosis of said patient.

6. The method according to claim 1 characterised in that the detection is carried out by

a) contacting said biological sample with an antibody immunoreactive with the PITX2 polypeptide to form an immunocomplex;
b) detecting said immunocomplex;
c) therefrom predicting the survival of said patient, characteristics of said cell proliferative disorder, and/or prognosis of said patient, and
d) comparing the quantity of said immunocomplex to the quantity of immunocomplex formed under identical conditions with the same antibody and a control sample from one or more patients with a known prognosis.

7. The method according to claim 1 characterised in that the detection is carried out by
a) contacting said biological sample with an antibody immunoreactive with the PITX2 polypeptide to form an immunocomplex;
b) detecting said immunocomplex; and
c) therefrom predicting the survival of said patient, characteristics of said cell proliferative disorder, and/or prognosis of said patient,
and wherein an increase in quantity of said immunocomplex in the sample from said subject relative to said control sample is indicative of a bad prognosis.

8. The method of claim 5, wherein said immunoassay is a radioimmunoassay or an ELISA or a Western blot.

9. The method of claim 1, wherein said detection is afforded by mRNA expression analysis.

10. The method of claim 9, comprising detecting the level of mRNA encoding a PITX2 polypeptide in a biological sample from a patient,

11. A method according to claim 10, wherein a increased concentration of said mRNA above the concentration determined for an individual known to have a good prognosis indicates a bad prognosis.

12. The method of claim 10, comprising the steps of:
(a) providing a polynucleotide probe which specifically hybridises or is identical to a polynucleotide consisting of SEQ ID NO: 19 or SEQ ID NO: 1,

(b) incubating said sample with said polynucleotide probe under high stringency conditions to form a specific hybridisation complex between a nucleic acid and said probe;

(c) detecting said hybridisation complex.

13. The method according to claim 12 wherein said nucleic acid is mRNA or a cDNA derived therefrom.

14. The method according to claim 13 wherein the detecting step further comprises the steps of:

a) producing a cDNA from mRNA in the sample;

b) providing two oligonucleotides which specifically hybridise to regions flanking a segment of the cDNA;

c) performing a polymerase chain reaction on the cDNA of step a) using the oligonucleotides of step b) as primers to amplify the cDNA segment; and

d) detecting the amplified cDNA segment.

15. Use of a polypeptide expressed from the PITX2 gene for differentiating or distinguishing between patients diagnosed with breast cancer, who have a good survival prognosis and patients who have a bad survival prognosis.

16. Use of a polypeptide expressed from the PITX2 gene for prediction of survival of a patient diagnosed with a cell proliferative disorder of the breast.

17. The method of claim 1 wherein said detection comprises determining the genetic parameters of the gene PITX2, its promoter and/or regulatory elements.

18. The method of claim 1 wherein said detection comprises determining the epigenetic parameters of the gene PITX2, its promoter and/or regulatory elements.

19. The method of claim 1, wherein said detection comprises determining the methylation status of one or more CpG positions of a target nucleic acid within the gene PITX2, its promoter and/or regulatory elements, in particular through the methylation analysis of a genomic DNA sequence according to SEQ ID NO: 1.

20. The method of claim 1, wherein said detection comprises determining the methylation status of one or more CpG positions of a target nucleic acid characterised as being identical to or hybridising under stringent or moderately stringent conditions to a sequence out of the group of SEQ ID NO: NOs 13, 18 and 19.

21. The method of claim 19, wherein the methylation analysis is afforded by contacting said target nucleic acid with one or more agents that convert cytosine bases that are unmethylated at the 5'-position thereof to a base that is detectably dissimilar to cytosine in terms of hybridisation properties.

22. The method of claim 21, wherein contacting said target nucleic acids with one or more agents comprises use of a solution selected from the group consisting of bisulfite, hydrogen sulfite, disulfite, and combinations thereof.

23. Use of a set of oligomer probes comprising at least two oligomers, in particular an oligonucleotide or peptide nucleic acid (PNA)-oligomer, said oligomer comprising in each case at least one base sequence having a length of at least 9 nucleotides which is complementary to, or hybridises under moderately stringent or stringent conditions to a pretreated genomic DNA according to one of the SEQ ID NO: 2 to SEQ ID NO: 5 and sequences complementary thereto, for detecting the cytosine methylation state and/or single nucleotide polymorphisms (SNPs) within one of the sequences according to SEQ ID NO: 1, and sequences complementary thereto.

24.A method for manufacturing an arrangement of different oligomers (array) fixed to a carrier material for analysing diseases associated with the methylation state of the CpG dinucleotides of one of SEQ ID NO: 1, and sequences complementary thereto wherein at least one oligomer, in particular an oligonucleotide or peptide nucleic acid (PNA)-oligomer, said oligomer comprising in each case at least one base sequence having a length of at least 9 nucleotides which is complementary to, or hybridises under moderately stringent or stringent conditions to a pretreated genomic DNA according to one of the SEQ ID NO: 2 to SEQ ID NO: 5 and sequences complementary thereto, is coupled to a solid phase.

25. A composition of matter comprising the following:

- a nucleic acid comprising a sequence at least 18 bases in length of a segment of the

chemically pretreated genomic DNA according to one of the sequences taken from the group comprising SEQ ID NO: 1 to SEQ ID NO: 5 and sequences complementary thereto, and

- a buffer comprising at least one of the following substances: 1 to 5 mM Magnesium Chloride, 100-500 µM dNTP, 0.5-5 units/10ul of taq polymerase, an oligomer, in particular an oligonucleotide or peptide nucleic acid (PNA)-oligomer, said oligomer comprising in each case at least one base sequence having a length of at least 9 nucleotides which is complementary to, or hybridises under moderately stringent or stringent conditions to a pretreated genomic DNA according to one of the SEQ ID NO: 2 to SEQ ID NO: 5 and sequences complementary thereto.

26. Use of the gene PITX2, its promoter and/or regulatory elements for predicting the survival of patients diagnosed with a cell proliferative disease.

27. Use of the mRNA of the gene PITX2 for predicting the survival of patients diagnosed with a cell proliferative disease.

28. A method for predicting the survival of patients diagnosed with a cell proliferative disease according to claim 19, comprising:

a) isolating or enriching genomic DNA from said biological sample ;

b) treating the genomic DNA, or a fragment thereof, with one or more reagents to convert 5-position unmethylated cytosine bases to uracil or to another base that is detectably dissimilar to cytosine in terms of hybridisation properties;

c) contacting the treated genomic DNA, or the treated fragment thereof, with an amplification enzyme and at least two primers comprising, in each case a contiguous sequence at least 18 nucleotides in length that is complementary to, or hybridises under moderately stringent or stringent conditions to a sequence selected from the group consisting of SEQ ID NO: 2 to 5, and complements thereof, wherein the treated DNA or a fragment thereof is either amplified to produce one or more amplificates, or is not amplified; and

d) determining, based on the presence or absence of, or on the quantity or on a property of said amplificate, the methylation state of at least one CpG dinucleotide sequence of SEQ ID NO: 1, or an average, or a value reflecting an average methylation state of a plurality of CpG dinucleotide sequences of SEQ ID NO: 1.

29. A method for detecting the survival of patients diagnosed with a cell proliferative disease of the breast according to claim 19, comprising the following steps of

a) obtaining, from a subject, a biological sample having subject genomic DNA;

b) treating the genomic DNA, or a fragment thereof, with one or more reagents to convert 5-position unmethylated cytosine bases to uracil or to another base that is detectably dissimilar to cytosine in terms of hybridisation properties;

c) amplifying one or more fragments of the treated DNA such that only DNA originating from breast or breast cell proliferative disorder cells are amplified

d) detecting the amplificates or characteristics thereof and thereby deducing on the survival of patients diagnosed with a cell proliferative disease of the breast.

30. Use of an oligomer, an oligonucleotide or peptide nucleic acid (PNA)-oligomer, said oligomer comprising in each case at least one base sequence having a length of at least 9 nucleotides which is complementary to, or hybridises under moderately stringent or stringent conditions to an artificially modified, chemically pretreated DNA according to one of the SEQ ID NO: 2 to SEQ ID NO: 5 and sequences complementary thereto, for differentiating or distinguishing between patients diagnosed with breast cancer, who have a good survival prognosis and patients who have a bad survival prognosis.

31. Use of a nucleic acid comprising a sequence of at least 18 bases in length of a segment of the artificially modified, chemically pretreated, DNA according to one of the sequences taken from the group comprising SEQ ID NO: 2 to SEQ ID NO: 5 and sequences complementary thereto, for differentiating or distinguishing between patients diagnosed with breast cancer, who have a good survival prognosis and patients who have a bad survival prognosis.

32. Use of an oligomer, an oligonucleotide or peptide nucleic acid (PNA)-oligomer, said oligomer comprising in each case at least one base sequence having a length of at least 9 nucleotides which is complementary to, or hybridises under moderately stringent or stringent conditions to an artificially modified, chemically pretreated DNA according to one of the SEQ ID NO: 2 to SEQ ID NO: 5 and sequences complementary thereto, for prediction of survival of a patient diagnosed with a cell proliferative disorder of the breast.

33. Use of a nucleic acid comprising a sequence of at least 18 bases in length of a segment of the artificially modified, chemically pretreated, DNA according to one of the sequences taken from the group comprising SEQ ID NO: 2 to SEQ ID NO: 5 and sequences complementary thereto, for prediction of survival of a patient diagnosed with a cell proliferative disorder of the breast.

34. Use of a nucleic acid represented by their SEQ ID NO: NO out of the group of nucleic acids according to SEQ ID NO: NOS : 6, 7, 8, 9, 10, 14, 15, 22, 23, 24, 25, 26, 27, 28, 29, 30 and 31, for differentiating or distinguishing between patients diagnosed with breast cancer, who have a good survival prognosis and patients who have a bad survival prognosis.

35. Use of a nucleic acid represented by their SEQ ID NO: NO out of the group of nucleic acids according to SEQ ID NO: NOS : 6, 7, 8, 9, 10, 14, 15, 22, 23, 24, 25, 26, 27, 28, 29, 30 and 31, for prediction of survival of a patient diagnosed with a cell proliferative disorder of the breast.

Abstract

The present invention relates to methods for predicting the survival of a human being diagnosed with a cell proliferative disorder of the breast tissues, characterised by a step of determining the expression level of PITX2 or the genetic or the epigenetic modifications of the genomic DNA associated with the gene PITX2. The invention also relates to sequences, oligonucleotides and antibodies which can be used within the described methods.

Figure 1

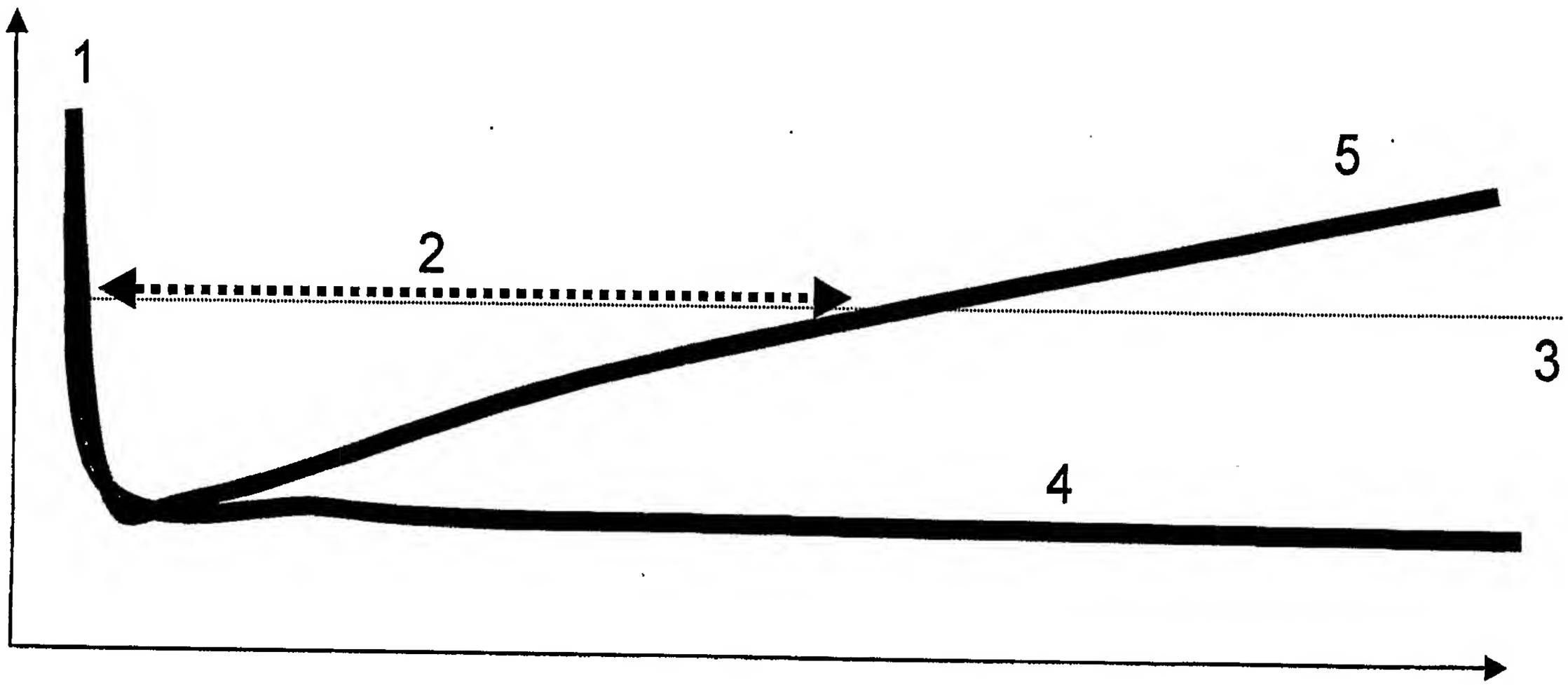

Figure 2

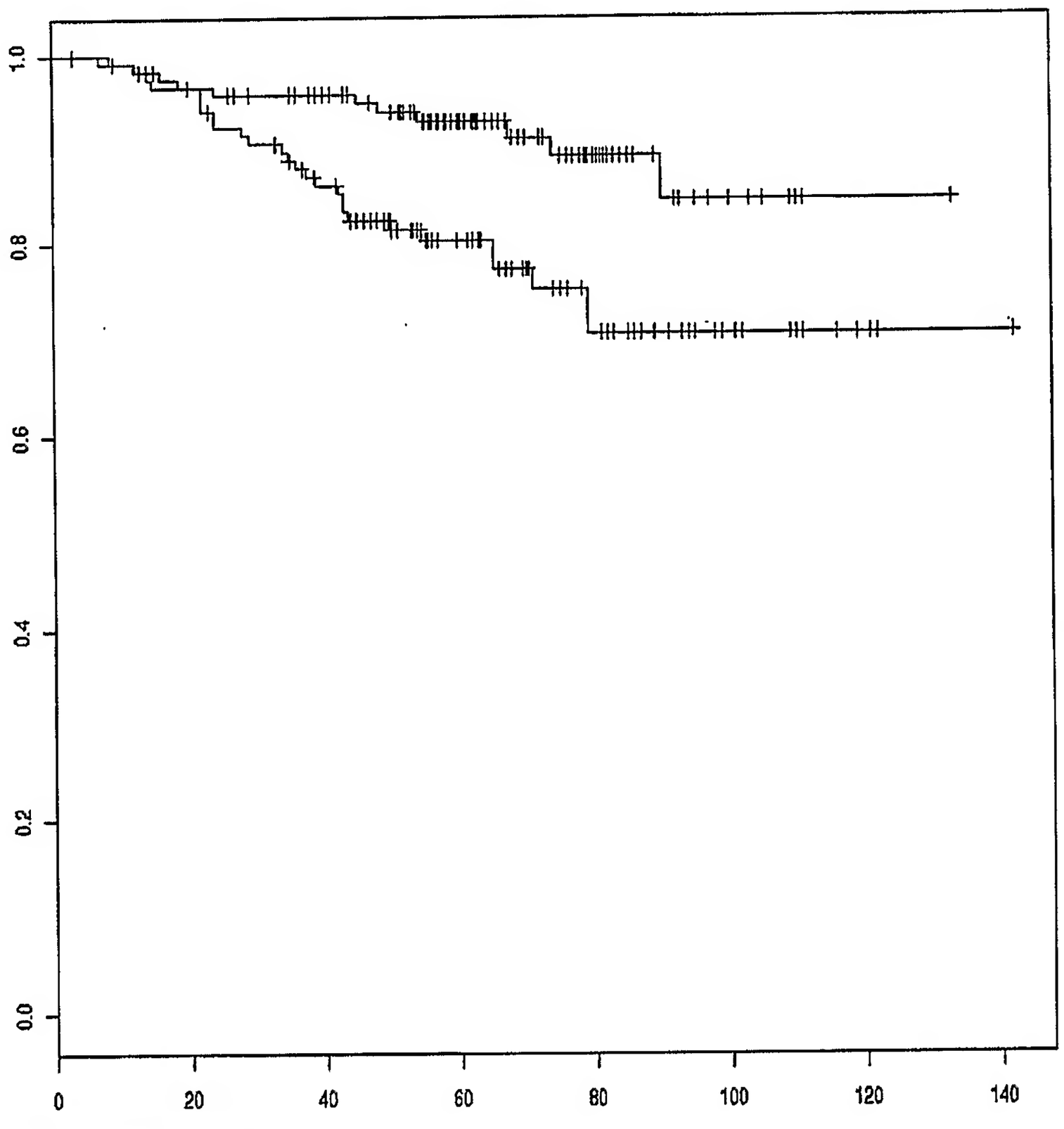

Figure 3

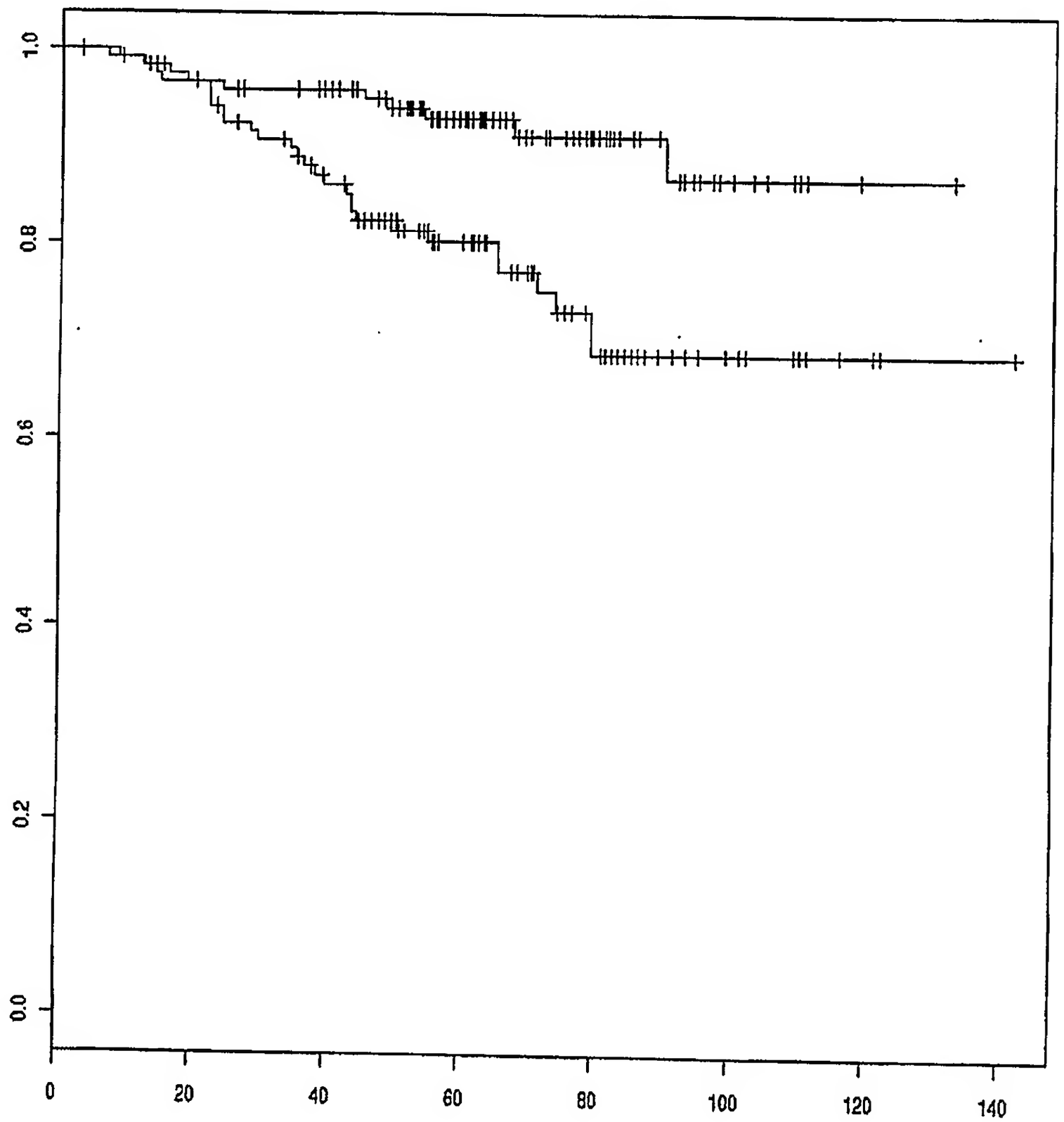

Figure 4

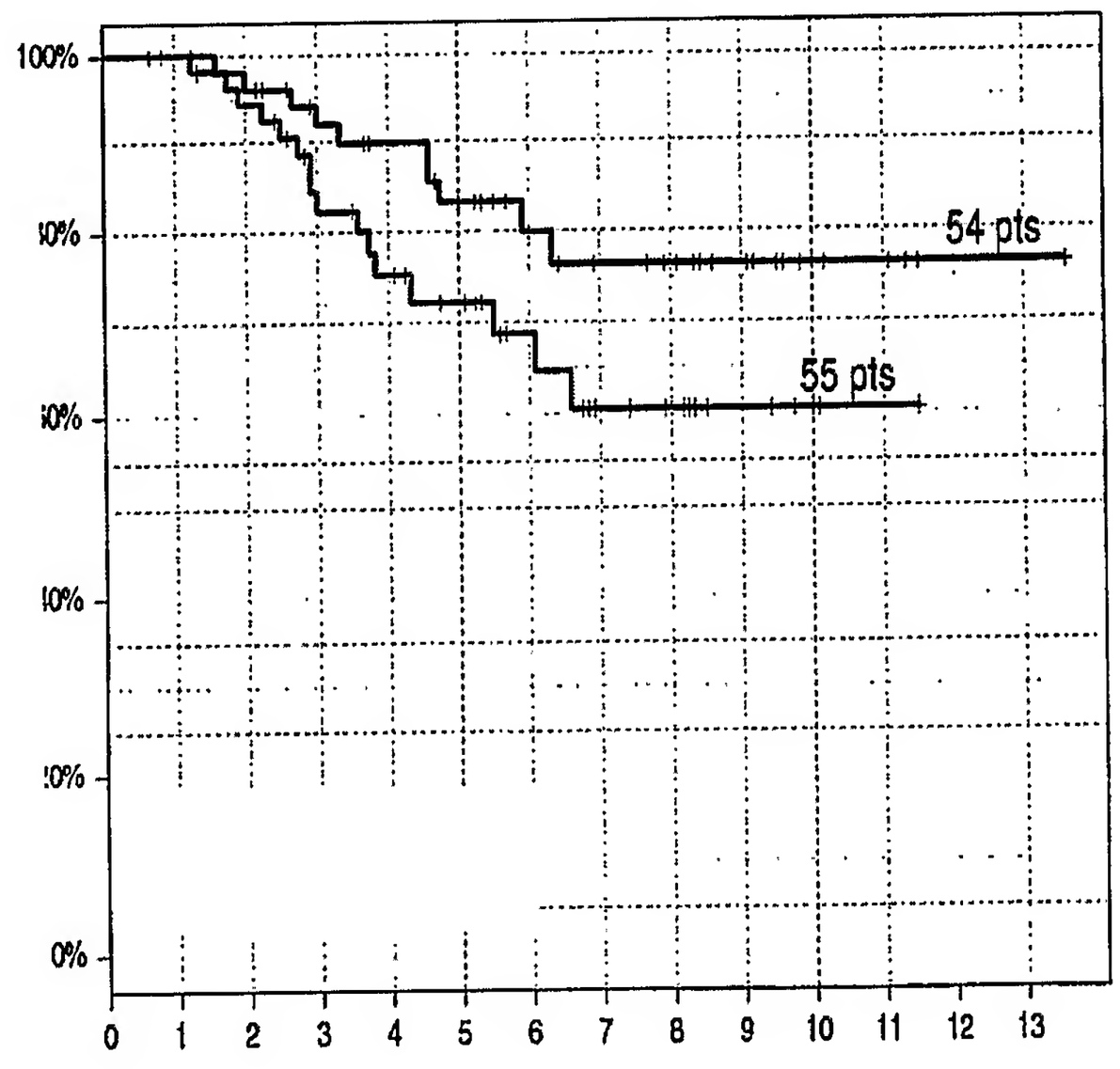

Figure 5

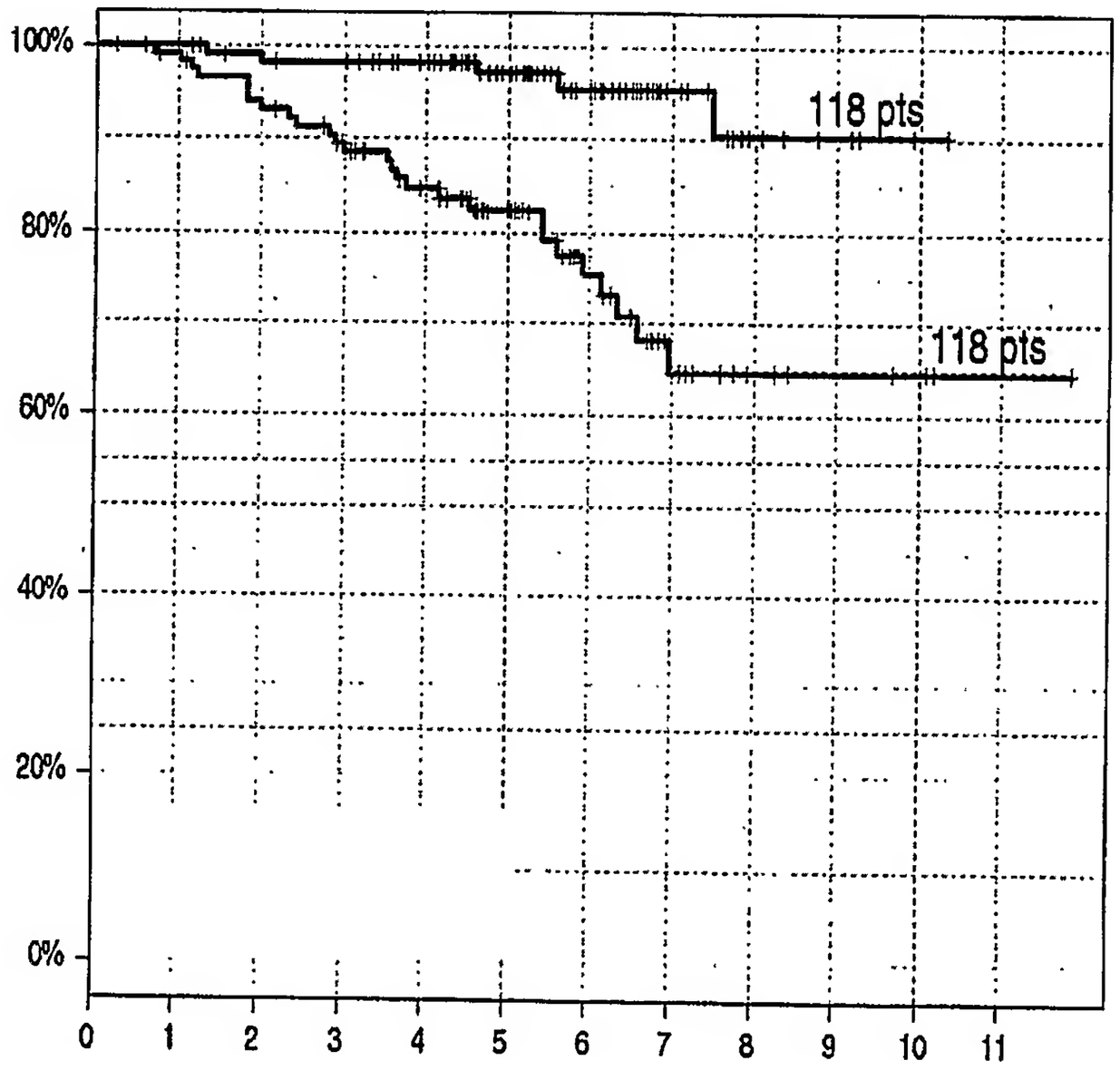

Figure 6

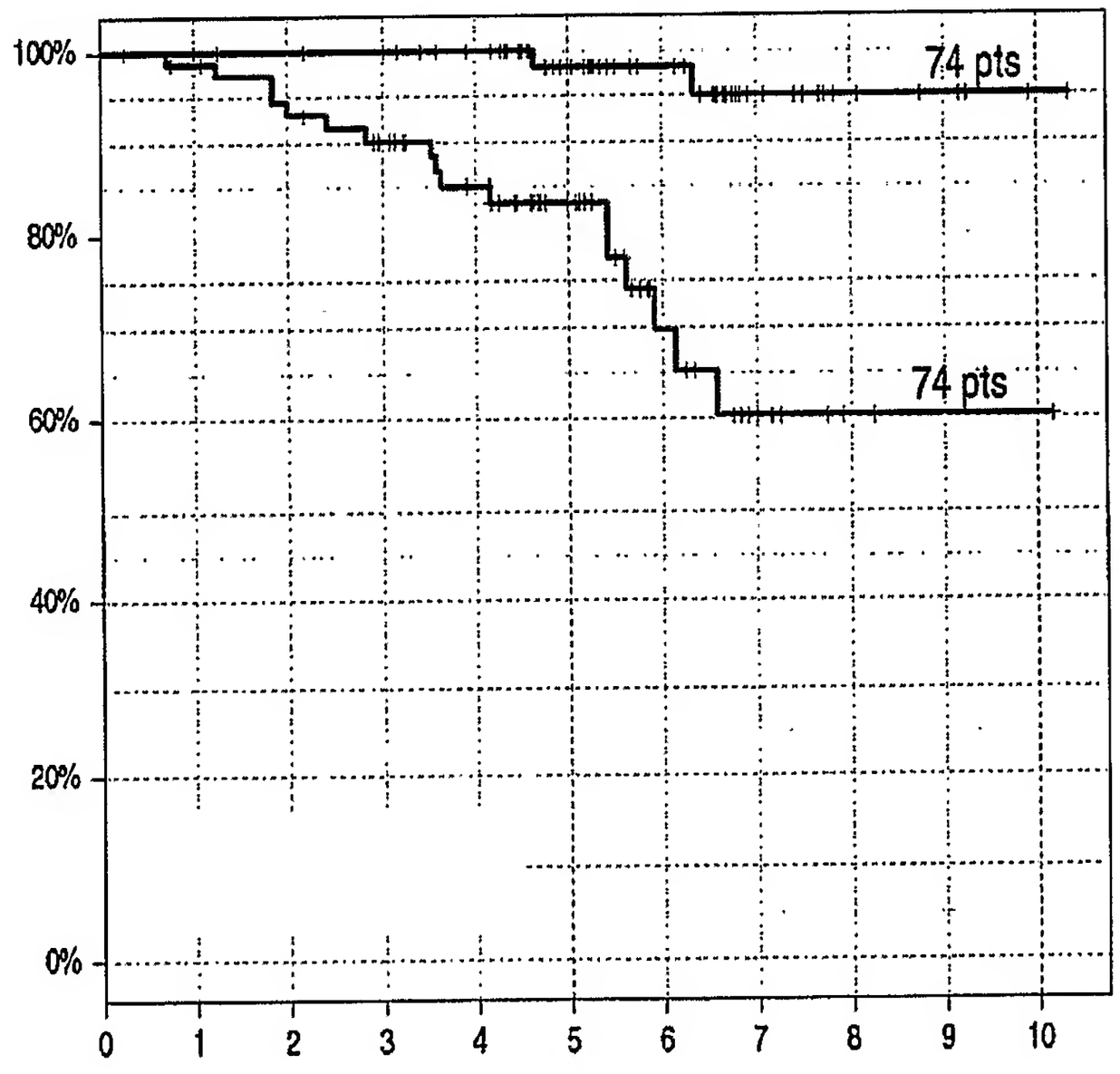

Figure 7

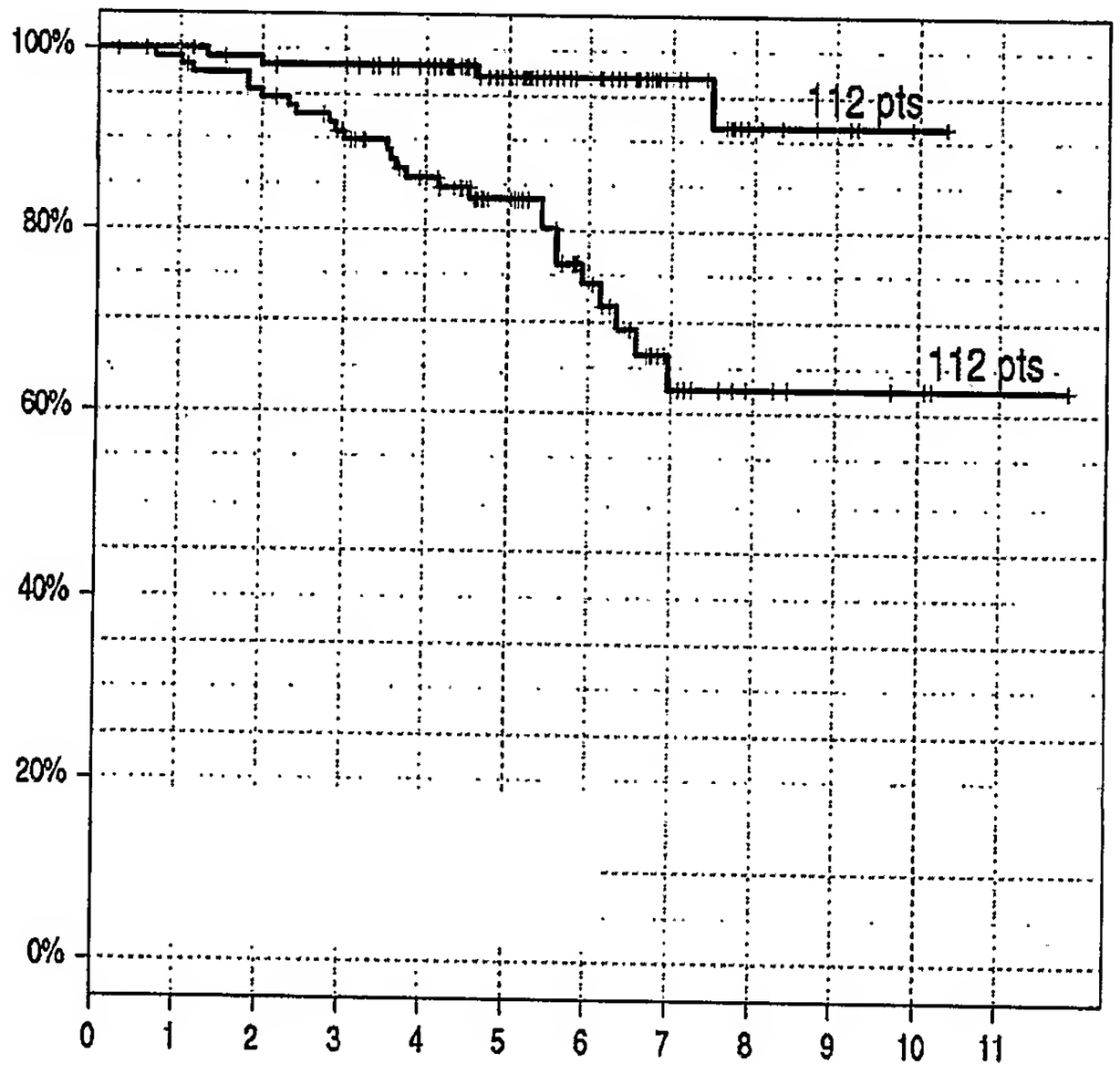

Figure 8

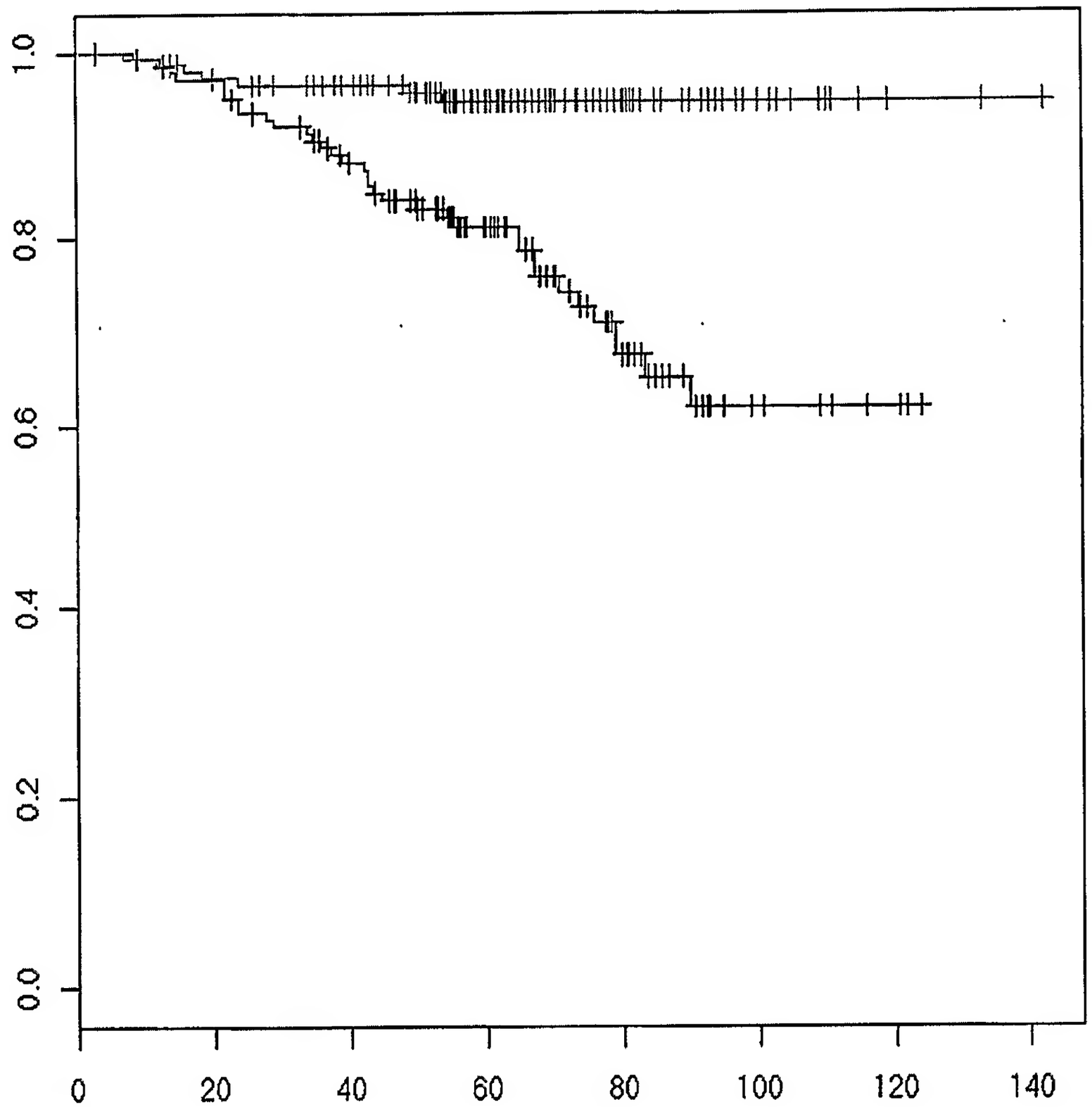

Figure 9

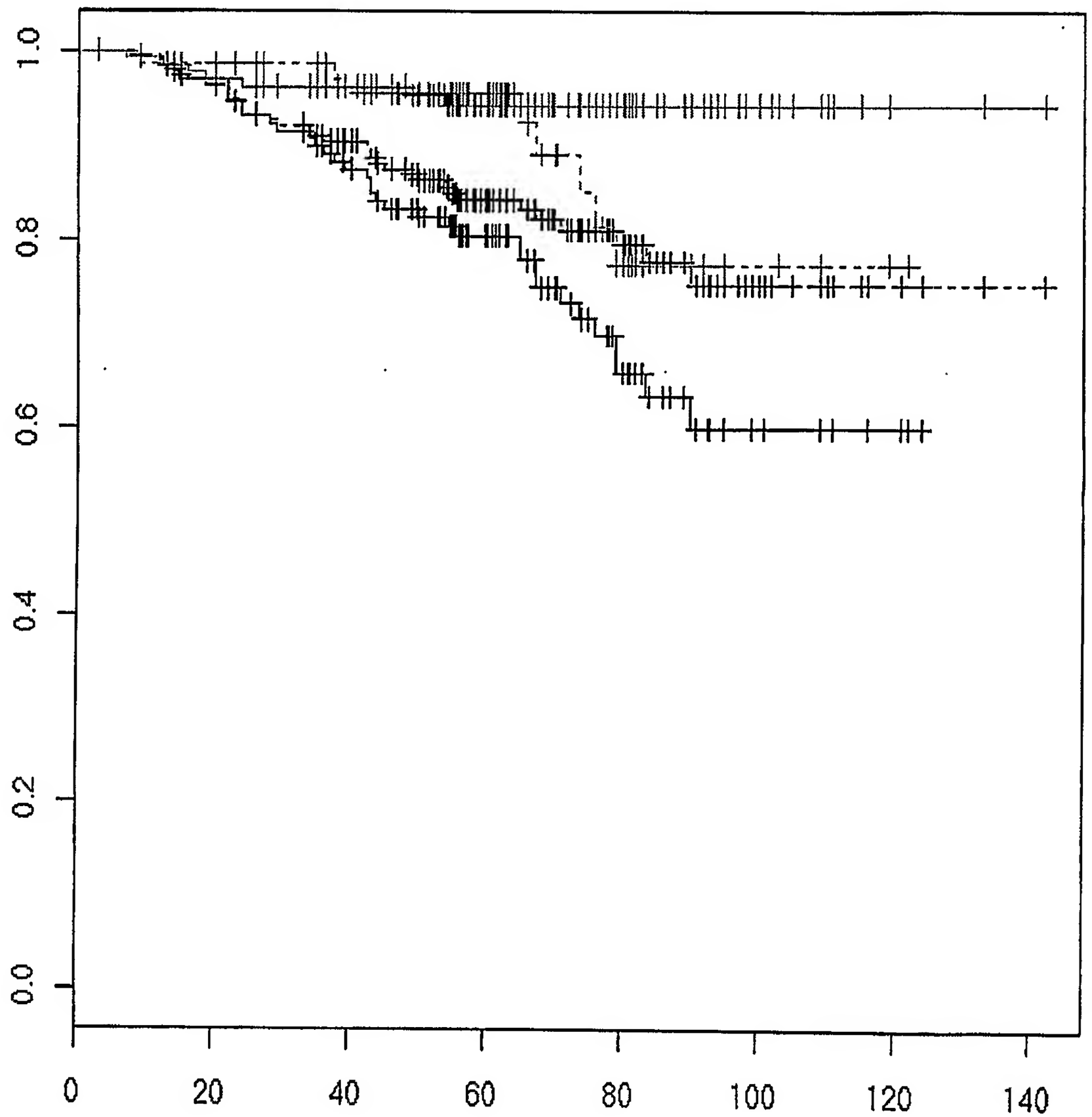

Figure 10

METNCRKLVSACVQLGVQPAAVECLFSKDSEIKKVEFTDSPESRKEAASSKFFPRQ
HPGANEKDKSQQGKNEDVGAEDPSKKKRQRRQRTHFTSQQLQELEATFQRNRYP
DMSTREEIAVWTNLTEARVRVWFKNRRAKWRKRERNQQAELCKNGFGPQFNGL
MQPYDDMYPGYSYNNWAAKGLTSASLSTKSFPFFNSMNVNPLSSQSMFSPPNSISS
MSMSSSMVPSAVTGVPGSSLNSLNNLNNLSSPSLNSAVPTPACPYAPPTPPYVYRDT
CNSSLASLRLKAKQHSSFGYASVQNPASNLSACQYAVDRPV

Sequence listing

<110> Epigenomics AG

<120> PITX2 - a marker to predict survival of patients diagnosed with breast cell proliferative disease

<160> 35

<210> 1
<211> 9001
<212> DNA
<213> Homo Sapiens

<400> 1

agctgatgga cttgctaaat ttctttcttc ttttttcttt ttcatattat ttgctagcca 60
taatggaatc ctctaggttt aagccaaaga aaaattggag agacaaaatt agattttgta 120
gccctttcc cccccgggaa tgccttttt tttctttta gtttctgatg aatggctatc 180
atttatttct accaaattta aataaggact gctgccttgt atgtttaact aggcaggcag 240
agggaactgg tttgtttagg aagcagtgac tgagatgtcc tggccaagtt agtgacagag 300
gaggggagaa agaatccaga ccaatttgta tgcagtatat tttactccca tgaaataaaa 360
cacatttgtt tcatatttgc tgaaaagtaa aacaataata ttgtacgaaa tgttatacac 420
agggtaggtt gtacatagca gtttcagaaa catcattgca tccaccagag aaactattct 480
aaaactgata ttcacacatt ttttataata ataataatat gttagaaaca tacagtgtgg 540
catttagtat atacactccc ttgctcgcaa gcgaaaaatc ctaatcgctt ctgtataaca 600
tgctttattt taaagcctaa cctttaaaaa cactgttgtg atattactaa caactgcttt 660
tataaaatta atttgacatt tcgatatata tacatccttt cagtcattta aatgttaaca 720
atgctaaact taaaaaataa caagcttata gtaatgttaa aatgtcatat ccagtcaaac 780
atttgtttgt gtatgtgtcc ttgcaactgt tagaaatact tgtagtgaaa gatgtcagac 840
actgaggaca tccctttgaa atcaaaggag ctctctcttt gattcagtgg tttccttttc 900
tctatatagc ttctctttct ctccctttct ttagtgccca cgaccttcta gcataattcc 960
cagtctttca agggcggagt tgccccatcc ggcaaggtcc taggatcccg gcgctgtggg 1020
tgcggctcac acgggccggt ccactgcata ctggcaagca ctcaggttgg aggccgggtt 1080
ctgcacgctg gcgtagccga agctggagtg ctgctttgct ttcagtctca ggctggccag 1140
gctcgagtta cacgtgtccc tataaacata cggaggagtc ggcggcgcgt aaggacaggc 1200
aggcgtcggc accgcggaat tcagcgacgg gctactcagg ttgttcaagt tattcaggct 1260
gttgagactg gagcccggga cgcctgtcac tgctgagggc accatgctgg acgacatgct 1320
catggacgag atagagttgg gtggggaaaa catgctctgt gatgacaggg ggttgacgtt 1380
catagagttg aagaagggga agctcttggt ggatagggag gcggatgtaa ggcccttggc 1440
ggcccagttg ttgtaggaat agcctgggta catgtcgtcg tagggctgca tgagcccatt 1500
gaactgcggc ccgaagccat tcttgcatag ctcggcctgc tggttgcgct ccctctttct 1560
ccatttggcc cgacgattct tgaaccaaac ctgggggcgg ttggggcaag ggagcaaaca 1620
gatgccacag tgcagattac taaaacttcc atcggaggcc aacccccgcc ttccccccgac 1680
acacacgcta gcgcactcac acaccctggc ctcgcttcac tgcaccgccc tgcacaccaa 1740
gataccaggg ccagctttca gttactggcc cgggtctcca ccaagcgcag gagacctggt 1800
ctgctctggc ctgcgagctg ggactcggag ctacgccaca aacctcagcc gaacgcatgg 1860
agacctgcgg acggtttgat cactcagcca ggcgtttctc caggtccaaa aacacttaat 1920
gtaaaacaaa cgcggggcag caggcttttc caacccttcc cggggcacct tgcaaacttg 1980
cttccattcc aaagccacag acccacggat gaggagaagg ggctggaagg gcactagagg 2040
atcgctcttt ctcccacgca attcctccct tccttccctg acctccactg tcgtccccca 2100
ccccctggta cgtgctccct taacagggac taggccgcca acactctttc tcgcctagca 2160
aaacaaccaa ataaagagca aaagaccacc tcttcgtcag ctcgttaact ccaggagctt 2220
ggcatattaa actccgggaa cccggaaagg gtagttttgg agattccccc ttctttcgct 2280
ctgcctcttc tttaccctaa gcccaccaca ggcctgtccg cgcgccaggc ccagccgggt 2340
cgtttggctt tgcaggcggc cacccaggcc ggccggcttc cacccgtgtc cggtggccca 2400
gccgcaaccc cgatcccaat ccacatcggg cctccctgtc gccccagacg gcggcttttg 2460
tgtattggag agaggcctgg cctgagatat ccgagctgac accagtgatg tttcacatta 2520
cacatctccg ccgggcccag ccgtgtaatc cgcttttttct cttttttcctt tcattcttga 2580
tttccttttt atccccccttc ctctttgcac ccgactgcta taaaaagcac gcctcactcc 2640
cacttggctc gacaagcagc cgccctggaa ggagaggcag ctgcaaggag agcccagcgc 2700
cgcggctaca aagcactagg gtggagctgc ggaatagcgg gcggggtggg agggcgtttt 2760
cgaaggatcc cagaaaaccc atagactctg tctttaatta cttgccattt ctaccctagg 2820

ccatctaaac tttgctcagg cgagaagagt acgtgagagg cccgttccct tgatgtgcaa 2880
gagagctaat gaaagactga ccttgctcaa aaccacgccg cccaggaccc agctctggct 2940
ctggacagtt aaactaaaac cattttcaac ttcttcccgg ccttttatcc accagcatag 3000
cctcatgcct tgcacaaatg ccacccagag agtgtcttca ttccctctga tttgggagag 3060
cattttggtc tttattcttt ttatcgttgt tttcttcttt ttgtttgctc tgctctaacc 3120
gggggcttta ttttttctac ccagagcact taatttttt tttttaacag caaagcctct 3180
ggatgccgct tgatttgctt gattctgttt tctgcttcca gaatcctaac aaatttggaa 3240
tcttccaccg accagcataa accaggacgt tgctattggg ttatttattt gagctcattt 3300
ttgccaatcc ataaagtaca gatttgctac aaagttaagg taagcccttt ttacaaaact 3360
atgattataa tttagaagag ggggtgtgag tttcaatttc cagagttcaa ctcctgagag 3420
aagataaata aaccaagcag aaaagtcttt cttctttttt tctttctcct tctaagagga 3480
ctagtagttg tgtattaaaa ctttgctccc ggagatcaca aaactaggaa atagggtgtg 3540
tgggagagac ctgaatggcc gaaacaaccg taaagaaggt gtaagaagcg cgagcccagg 3600
agggaaaaag ctgggccagg gccgggacaa aggtttccca gggagggcca actcttccgt 3660
gtctctggcg ggttttcctt gttaaaggct cacaggttgg agcctgttcg cggctcttgg 3720
cctggtaggg attttattag ctctgctctg gcaactgcaa gccaggaaca caatgtcctg 3780
tgcaggggat tgcccatgca gcccagctcg tgagatcgcg ggatggcggg gcagtgagcc 3840
ggtgccgctc tgggagcctg agccagggcg gcagtcctgt cggcctcgga gagggaactg 3900
taatctcgca accaggccgc cgcgaggcct tctgcctttg caaagctgcg ccccaccggc 3960
gccctcccag gcggcgctgc cttccacatt ctctcctggt ctacttggcc tgtacctcca 4020
caacatcctc cccccatccc tcccagactc cgtgctggct cctacccgga ctcgggcttc 4080
cgtaaggttg gtccacacag cgatttcttc gcgtgtggac atgtccgggt agcggttcct 4140
ctggaaagtg gcctccagct cctggagctg ctggctggta aagtgagtcc gctgccgcct 4200
ttgccgcttc ttcttagacg ggtcctcggc gcccacgtcc tcattcttcc cctgctggct 4260
tttatctttc tctgaaaacg aaacacacac actttcccgt cagcatgccc acctgcaacg 4320
cggacgccaa ctggaccggc ggcagaagcc gtggaagagc tgggctgcct ggcgccggag 4380
gagggtgcgc gcggcggctc cgggccgcga ggagcgctgc gcctgtgggg tgtgcaggcg 4440
caagtgtggg tgtccgcgcc ccatttcctc ccctcccccа gcgccgcacg ttttatttac 4500
atgtttatct cactgcagcg gcacattcac ttttatagcc tgtgctttca agtatattta 4560
tacacctctg cgcagacaca ccaaatctcc tgggacgcgc acacgcgcgt ggtttacaga 4620
ccccccctccc cctcgcagaa agctcagatt tccatgcggt ttgggaaggc taggaaaaga 4680
tgtggggatt cggttgggca ccgaagttcg ccggcccttt cccaaaaaaa aaaaaaaaat 4740
gcctcttcgc gaagggcatt tctgagtggt ttcaggcaat ttcctaacga gtggagctcc 4800
tcgggagctg aaagccgaga ggaaaacagg gacagaggtc ggcggcctct gaaggtcctc 4860
gaatcaagat gctgggattt ttgtgaccca ggaaacagaa gggaggccag ggtacgaata 4920
gagagggcgg cagaattgct cgcgcccctta gcgccccagg agccgggccg gtcgagggag 4980
aactaaaggg atgcggggta gtcaaaattc cggctcccgg aagttctgcg gggagccagg 5040
cgaacgacca ctcccaccac gcctcccccc ggaggggctg acttccttgg ggcgagaggg 5100
agcgggtggc gcagagcagc tgagcgggaa tgtctgcagg gcggcgcggc gccttacctg 5160
cggcctccgg gctggaggtg tcggagatgg tgtgcacctc cagcctgtgc ttggaggagt 5220
ccagcgaccg gggctgaccg ggagccagaa ccgaagccat ggctaacggc tggggatggt 5280
gacaggaaga tgaggagacg gccgacagct tggtccccgc tgctcggtgc tccaagtgaa 5340
gcgggccttt catgcagttc atggacgagg gagcgcgacg ctctactagt ccttggctac 5400
tgccccgccg agcccccgta gccgccgctg cccgctccgg gtcgcgctct aggcgcggag 5460
tttccccgct gcggggagag ccaggggacg caacccccgc cgagttctca agccaagctg 5520
cccccgtctc ctccggaagg ctcaagcgaa aaagtccgga gacggaaagt cagcgggcaa 5580
acgaagacat gggatgtggg cagaagggca ccactcagag cgtctttagg gagcaggctt 5640
ccaagctcca aagcgaaaca agagtgggca aagacccccт tcttctctcc ctcccctccc 5700
caagaacccc tccaataagg aaagctaacg ccgaccgcgc tctgcccgcc ccccccccac 5760
gcggcagccc tgacagagaa gtgtcaagag tgacagggac aggtaggtga tattagatcc 5820
cctgcggcgg cagcagccgc tgcagccacg acgcggccct ctgagcgcac cctccgcaac 5880
gcgcacacgc acacccctcg ggcggtcgaa caggagccgg gccttgccgc agctcagctc 5940
caggcaccca ggcgagcgac ggaccagatc tgcggctccg cgcttccctg ttggcctaac 6000
atcttaaaac cagaggcggg cttcctggtg ccgagacgtc actccgccgc ggccctcccc 6060
agccctctcc gcctccgcct cctcccagac ccttctccgg gtgcgactga cgtggctccg 6120
caccaatcag gacgccccga gccgcggtgg agggactgtc ctgcctgcac ctatcagcag 6180
tgcggggccg ggctactgcc tcgccgtgcg cactgggtct acacaggcaa gctcccggga 6240
attcagctcc tgcccagccc aaggcgatcc ggcttttagt acgaacccaa aggtgaagag 6300
atgaggctag gagtcgaagg cttgggagaa gagagtggaa tggtcaagaa gagaaaggta 6360
caaggatcaa caagacaccc actctttgtg tctcactaca tccatttcca atcccccacc 6420
ccatataaaa aggagacacg ttacttaaaa ctagaaaatt tgaaaaacag caacaaatca 6480
cctctccgat cttaaatttt ccaaacagcc tgtcaagtga atgctgcgct aatctgaaga 6540
agctttaatt gcaaagaaga cagagccctg aaaaggcagg ctaataaatt agaaatcgag 6600

aagcaaatgg acccgtcaaa agaaaattac cttgacttta aacgaacaac tgtttggtgg 6660
ttcactctgg atttatacaa gaataaaaag tcgcctcaga tcacgttctc tgtgatgctt 6720
attagtcccc agacagaaaa cacacaatag aagagaaacc ctaacccagc gttttcaaaa 6780
tgctgaaagc ttatccattc tacttaacgt tgattaagac acatatccta gatctttcaa 6840
attccttgta cactgtatta agctcgtcct aacccgagag agccacgctt taaattcgac 6900
tctcttgttt actttattat caatcagatt taaatccata aagcctgtag aatcaacaac 6960
cttgagctaa ttatatatga aatatgcctt aatgaatttc catacaatta agaatgttgc 7020
caaataacca atttcaagga taatttttaa cagtcatttt cttttcccag tgagctcaag 7080
gctgtcttga gccattaaag tccaagcagg cagaaggggt gtgtgtgagc taagggcgaa 7140
aagcctagaa ctgcgctcaa ctagcaaaag caaaacctta tttatataaa acaaaaaaaa 7200
tcacctttgg agacatcaac tctttatagc actgtttcca agcaaattta atttccaaag 7260
aaattaaaga aagaaatcca aacatattca aaataatttt tgaaagtcct tttgtccccc 7320
agcataggtc agctggagag gacaaactaa tctcctctgg gtttctgcat gggcgattgt 7380
tttactatgg agttagtgtt atcatctctg aatgtgtatt tgtttgacat tacagtcaat 7440
gatttgcaat gccagcatga agtatcttta aaacactccc tccttgtcct tgttcacaag 7500
attgggaaac ttattcgatg tggaacaaag tggatgaagc agactacaaa tatatttgca 7560
acttatgtgt ctctcctttg ccctgaccac ccccaaaccc tatctgcaac tcctccccat 7620
tttaaacttg cagtccaaag acgcacatga gaattgtttt tcagtctttc ttcaccagta 7680
tcatcccact ttaagaataa tttagctgca agggaggaat ttcttcatag taagctttaa 7740
atcagcattt ctgcttttaa ccttttattc cactttaccc cattccacac atacagacac 7800
ctgctcagag taaaacacat cctcatgtga caggtctgca ttagctgagg ctcatacatc 7860
cagctatatt aggtcctgca atcttatcac taaattatac acattacact agcagcctgt 7920
tggtaaagaa ggttaaatta atttacattc tgctcattat ctggtgctta aatgacgcat 7980
tttatcccgg agatttggcg gagaatctcc ttctcagacc ccacagcgtt tcactgaaga 8040
caatgctctt acatttgtag tggtttttaa tctgataaga ctctaatttg cttaagtctt 8100
ttaaataagg gttttaaatg tttctagccg ttttcttatt gaatttcctc taattccccc 8160
aagatcataa agtatatgtg taaagtaaat atttcctccc attgcactgc cagccgatga 8220
cctataacta agtcaataag aatccagctc ttttctgctg aatgtgttta ctaatcatat 8280
tccagtttct tcttttaaac ctcagaatag ctgtggtccc cacaatacca tgccccttaa 8340
agcttcattt catgaaggga ctccatcaca ttaaagaatg aaaaaaatct ccactgtagt 8400
tagtatacac agcccctcac tccttgtttt tcaagattca aaccccagag ctgcaaatat 8460
ttttggaagc ttgggtgtta atgccttatt ttagaaagcc gagaagcccc acagagccat 8520
atagatttct aaacccatct ctcataaacc cacagaattt tgataaaagc tctggtggct 8580
ccactctacc gatggaactt tcatcacgac aaatatacat gtatgaagga cctcaatcag 8640
cctccaaagt ggttgaaaaa cccaagggca cgtgactgct cctcatagtg ccaacgtgtg 8700
cgagatgttg gaagcactgg ggatcagcag cagcctagat gcctaaaaag ataaggtgtc 8760
ctaatttgtg tggacccatt gaagtcaagt ggtgaataaa gacaattatc tagataattc 8820
agattaaagt aaaagcaaaa ccatatctat ttgtatatat atattcacat ccattttata 8880
ttacagacac acacacgcac acacacactg gctctgtaaa caactgactc aaagtgagga 8940
ttttctttgc atttttccag caggagtttc aacattctcc taatctccta atcactttac 9000
a 9001

<210> 2
<211> 9001
<212> DNA
<213> Artificial Sequence

<220>
<223> chemically treated genomic DNA (Homo sapiens)

<400> 2

agttgatgga tttgttaaat ttttttttt ttttttttt tttatattat ttgttagtta 60
taatggaatt ttttaggttt aagttaaaga aaaattggag agataaaatt agattttgta 120
gttttttttt ttttcgggaa tgttttttt ttttttttta gtttttgatg aatggttatt 180
atttatttt attaaattta aataaggatt gttgttttgt atgtttaatt aggtaggtag 240
agggaattgg tttgtttagg aagtagtgat tgagatgttt tggttaagtt agtgatagag 300
gaggggagaa agaatttaga ttaatttgta tgtagtatat tttatttta tgaaataaaa 360
tatatttgtt ttatatttgt tgaaaagtaa aataataata ttgtacgaaa tgttatatat 420
agggtaggtt gtatatagta gttttagaaa tattattgta tttattagag aaattatttt 480
aaaattgata tttatatatt ttttataata ataataatat gttagaaata tatagtgtgg 540
tatttagtat atatattttt ttgttcgtaa gcgaaaaatt ttaatcgttt ttgtataata 600
tgttttattt taaagtttaa tttttaaaaa tattgttgtg atattattaa taattgtttt 660

tataaaatta atttgatatt tcgatatata tatatttttt tagttattta aatgttaata 720
atgttaaatt taaaaaataa taagtttata gtaatgttaa aatgttatat ttagttaaat 780
atttgtttgt gtatgtgttt ttgtaattgt tagaaatatt tgtagtgaaa gatgttagat 840
attgaggata ttttttttgaa attaaaggag ttttttttttt gatttagtgg ttttttttttt 900
tttatatagt ttttttttttt ttttttttttt ttagtgttta cgatttttta gtataatttt 960
tagtttttta agggcggagt tgttttattc ggtaaggttt taggatttcg gcgttgtggg 1020
tgcggtttat acgggtcggt ttattgtata ttggtaagta tttaggttgg aggtcgggtt 1080
ttgtacgttg gcgtagtcga agttggagtg ttgttttgtt tttagtttta ggttggttag 1140
gttcgagtta tacgtgtttt tataaatata cggaggagtc ggcggcgcgt aaggataggt 1200
aggcgtcggt atcgcggaat ttagcgacgg gttatttagg ttgtttaagt tatttaggtt 1260
gttgagattg gagttcggga cgtttgttat tgttgagggt attatgttgg acgatatgtt 1320
tatggacgag atagagttgg gtggggaaaa tatgttttgt gatgataggg ggttgacgtt 1380
tatagagttg aagaagggga agtttttggt ggatagggag gcggatgtaa ggtttttggc 1440
ggtttagttg ttgtaggaat agtttgggta tatgtcgtcg tagggttgta tgagtttatt 1500
gaattgcggt tcgaagttat ttttgtatag ttcggtttgt tggttgcgtt ttttttttttt 1560
ttatttggtt cgacgatttt tgaattaaat ttgggggcgg ttggggtaag ggagtaaata 1620
gatgttatag tgtagattat taaaattttt atcggaggtt aattttcgtt ttttttcgat 1680
atatacgtta gcgtatttat atattttggt ttcgttttat tgtatcgttt tgtatattaa 1740
gatattaggg ttagttttta gttattggtt cgggttttta ttaagcgtag gagatttggt 1800
ttgttttggt ttgcgagttg ggattcggag ttacgttata aattttagtc gaacgtatgg 1860
agatttgcgg acggtttgat tatttagtta ggcgtttttt taggtttaaa aatatttaat 1920
gtaaaataaa cgcggggtag taggtttttt taatttttttt cggggtattt tgtaaatttg 1980
tttttatttt aaagttatag atttacggat gaggagaagg ggttggaagg gtattagagg 2040
atcgtttttt tttttacgta atttttttttt ttttttttttg atttttattg tcgtttttta 2100
ttttttggta cgtgtttttt taatagggat taggtcgtta atatttttttt tcgtttagta 2160
aaataattaa ataaagagta aaagattatt ttttcgttag ttcgttaatt ttaggagttt 2220
ggtatattaa atttcgggaa ttcggaaagg gtagttttgg agatttttttt ttttttcgtt 2280
ttgtttttttt tttattttaa gtttattata ggtttgttcg cgcgttaggt ttagtcgggt 2340
cgtttggttt tgtaggcggt tatttaggtc ggtcggtttt tattcgtgtt cggtggttta 2400
gtcgtaattt cgattttaat ttatatcggg ttttttttgtc gttttagacg gcggtttttg 2460
tgtattggag agaggtttgg tttgagatat tcgagttgat attagtgatg ttttatatta 2520
tatattttcg tcgggtttag tcgtgtaatt cgtttttttt ttttttttttt ttatttttga 2580
ttttttttttt atttttttttt ttttttgtat tcgattgtta taaaaagtac gttttatttt 2640
tatttggttc gataagtagt cgttttggaa ggagaggtag ttgtaaggag agtttagcgt 2700
cgcggttata aagtattagg gtggagttgc ggaatagcgg gcggggtggg agggcgtttt 2760
cgaaggattt tagaaaattt atagattttg tttttaatta tttgttattt ttattttagg 2820
ttatttaaat tttgtttagg cgagaagagt acgtgagagg ttcgtttttt tgatgtgtaa 2880
gagagttaat gaaagattga ttttgtttaa aattacgtcg tttaggattt agttttggtt 2940
ttggatagtt aaattaaaat tatttttaat ttttttttcgg ttttttattt attagtatag 3000
ttttatgttt tgtataaatg ttatttagag agtgttttta tttttttttga tttgggagag 3060
tattttggtt tttatttttt ttatcgttgt ttttttttttt ttgtttgttt tgttttaatc 3120
gggggttttta ttttttttat ttagagtatt taatttttttt ttttttaatag taaagttttt 3180
ggatgtcgtt tgatttgttt gattttgttt tttgttttta gaattttaat aaatttggaa 3240
ttttttatcg attagtataa attaggacgt tgttattggg ttatttattt gagtttattt 3300
ttgttaattt ataaagtata gatttgttat aaagttaagg taagtttttt ttataaaatt 3360
atgattataa tttagaagag ggggtgtgag ttttaatttt tagagtttaa tttttgagag 3420
aagataaata aattaagtag aaaagttttt ttttttttttt ttttttttttt tttaagagga 3480
ttagtagttg tgtattaaaa ttttgttttc ggagattata aaattaggaa atagggtgtg 3540
tgggagagat ttgaatggtc gaaataatcg taaagaaggt gtaagaagcg cgagtttagg 3600
agggaaaaag ttgggttagg gtcgggataa aggtttttta gggagggtta atttttttcgt 3660
gtttttggcg ggtttttttt gttaaaggtt tataggttgg agtttgttcg cggtttttgg 3720
tttggtaggg attttattag ttttgtttttg gtaattgtaa gttaggaata taatgttttg 3780
tgtaggggat tgtttatgta gtttagttcg tgagatcgcg ggatggcggg gtagtgagtc 3840
ggtgtcgttt tgggagtttg agttagggcg gtagttttgt cggtttcgga gagggaattg 3900
taatttcgta attaggtcgt cgcgaggttt tttgtttttg taaagttgcg ttttatcggc 3960
gtttttttag gcggcgttgt ttttttatatt ttttttttggt ttatttggtt tgtattttta 4020
taatattttt ttttttatttt ttttagattt cgtgttggtt tttattcgga ttcgggtttt 4080
cgtaaggttg gtttatatag cgatttttttc gcgtgtggat atgttcgggt agcggttttt 4140
ttggaaagtg gtttttagtt tttggagttg ttggttggta aagtgagttc gttgtcgttt 4200
ttgtcgtttt tttttagacg ggttttcggc gtttacgttt ttatttttttt tttgttggtt 4260
tttatttttt tttgaaaacg aaatatatat atttttttcgt tagtatgttt atttgtaacg 4320
cggacgttaa ttggatcggc ggtagaagtc gtggaagagt tgggttgttt ggcgtcggag 4380
gagggtgcgc gcggcggttt cgggtcgcga ggagcgttgc gtttgtgggg tgtgtaggcg 4440

taagtgtggg tgttcgcgtt ttatttttt ttttttttta gcgtcgtacg ttttatttat 4500
atgtttattt tattgtagcg gtatatttat ttttatagtt tgtgttttta agtatattta 4560
tatatttttg cgtagatata ttaaattttt tgggacgcgt atacgcgcgt ggtttataga 4620
ttttttttt tttcgtagaa agtttagatt tttatgcggt ttgggaaggt taggaaaaga 4680
tgtggggatt cggttgggta tcgaagttcg tcggtttttt tttaaaaaaa aaaaaaaaat 4740
gtttttttcgc gaagggtatt tttgagtggt tttaggtaat tttttaacga gtggagtttt 4800
tcgggagttg aaagtcgaga ggaaaatagg gatagaggtc ggcggttttt gaaggttttc 4860
gaattaagat gttgggattt ttgtgattta ggaaatagaa gggaggttag ggtacgaata 4920
gagagggcgg tagaattgtt cgcgttttta gcgttttagg agtcgggtcg gtcgagggag 4980
aattaaaggg atgcggggta gttaaaattt cggttttcgg aagttttgcg gggagttagg 5040
cgaacgatta tttttattac gtttttttttc ggaggggttg attttttttgg ggcgagaggg 5100
agcgggtggc gtagagtagt tgagcgggaa tgtttgtagg gcggcgcggc gttttatttg 5160
cggttttcgg gttggaggtg tcggagatgg tgtgtatttt tagtttgtgt ttggaggagt 5220
ttagcgatcg gggttgatcg ggagttagaa tcgaagttat ggttaacggt tggggatggt 5280
gataggaaga tgaggagacg gtcgatagtt tggttttcgt tgttcggtgt tttaagtgaa 5340
gcgggttttt tatgtagttt atggacgagg gagcgcgacg ttttattagt ttttggttat 5400
tgtttcgtcg agttttcgta gtcgtcgttg ttcgtttcgg gtcgcgtttt aggcgcggag 5460
ttttttcgtt gcggggagag ttaggggacg taattttcgt cgagttttta agttaagttg 5520
ttttcgtttt tttcggaagg tttaagcgaa aaagttcgga gacggaaagt tagcgggtaa 5580
acgaagatat gggatgtggg tagaagggta ttatttagag cgtttttagg gagtaggttt 5640
ttaagtttta aagcgaaata agagtgggta aagatttttt ttttttttttt ttttttttttt 5700
taagaatttt tttaataagg aaagttaacg tcgatcgcgt tttgttcgtt ttttttttac 5760
gcggtagttt tgatagagaa gtgttaagag tgatagggat aggtaggtga tattagattt 5820
tttgcggcgg tagtagtcgt tgtagttacg acgcggtttt ttgagcgtat ttttcgtaac 5880
gcgtatacgt atatttttcg ggcggtcgaa taggagtcgg gttttgtcgt agtttagttt 5940
taggtattta ggcgagcgac ggattagatt tgcggtttcg cgttttttttg ttggtttaat 6000
attttaaaat tagaggcggg ttttttggtg tcgagacgtt atttcgtcgc ggtttttttt 6060
agtttttttc gttttcgttt ttttttagat ttttttttcgg gtgcgattga cgtggtttcg 6120
tattaattag gacgtttcga gtcgcggtgg agggattgtt ttgtttgtat ttattagtag 6180
tgcggggtcg ggttattgtt tcgtcgtgcg tattgggttt atataggtaa gttttcggga 6240
atttagtttt tgtttagttt aaggcgattc ggtttttagt acgaatttaa aggtgaagag 6300
atgaggttag gagtcgaagg tttgggagaa gagagtggaa tggttaagaa gagaaaggta 6360
taaggattaa taagatattt atttttttgtg ttttattata tttatttttta attttttatt 6420
ttatataaaa aggagatacg ttatttaaaa ttagaaaatt tgaaaaatag taataaatta 6480
ttttttcgat tttaaatttt ttaaatagtt tgttaagtga atgttgcgtt aatttgaaga 6540
agttttaatt gtaaagaaga tagagttttg aaaaggtagg ttaataaatt agaaatcgag 6600
aagtaaatgg attcgttaaa agaaaattat tttgatttta aacgaataat tgtttggtgg 6660
tttattttgg atttatataa gaataaaaag tcgttttaga ttacgttttt tgtgatgttt 6720
attagttttt agatagaaaa tatataatag aagagaaatt ttaatttagc gtttttaaaa 6780
tgttgaaagt ttatttattt tatttaacgt tgattaagat atatatttta gatttttaa 6840
atttttgta tattgtatta agttcgtttt aattcgagag agttacgttt taaattcgat 6900
ttttttgttt attttattat taattagatt taaatttata aagtttgtag aattaataat 6960
tttgagttaa ttatatatga aatatgtttt aatgaatttt tatataatta agaatgttgt 7020
taaataatta attttaagga taatttttaa tagttatttt ttttttttag tgagtttaag 7080
gttgttttga gttattaaag tttaagtagg tagaaggggt gtgtgtgagt taagggcgaa 7140
aagtttagaa ttgcgtttaa ttagtaaaag taaaatttta tttatataaa ataaaaaaaa 7200
ttatttttgg agatattaat tttttatagt attgttttta agtaaattta atttttaaag 7260
aaattaaaga aagaaattta aatatattta aaataatttt tgaaagtttt tttgtttttt 7320
agtataggtt agttggagag gataaattaa ttttttttgg gtttttgtat gggcgattgt 7380
tttattatgg agttagtgtt attatttttg aatgtgtatt tgtttgatat tatagttaat 7440
gatttgtaat gttagtatga agtatttttа aaatattttt tttttgtttt tgtttataag 7500
attgggaaat ttattcgatg tggaataaag tggatgaagt agattataaa tatatttgta 7560
atttatgtgt ttttttttttg ttttgattat ttttaaattt tatttgtaat ttttttttat 7620
tttaaatttg tagtttaaag acgtatatga gaattgtttt ttagtttttt tttattagta 7680
ttatttttatt ttaagaataa tttagttgta agggaggaat ttttttatag taagttttaa 7740
attagtattt ttgtttttaa ttttttattt tattttattt tattttatat atatagatat 7800
ttgtttagag taaaatatat ttttatgtga taggtttgta ttagttgagg tttatatatt 7860
tagttatatt aggttttgta attttattat taaattatat atattatatt agtagtttgt 7920
tggtaaagaa ggttaaatta atttatattt tgtttattat ttggtgttta aatgacgtat 7980
tttatttcgg agatttggcg gagaattttt tttttagatt ttatagcgtt ttattgaaga 8040
taatgttttt atatttgtag tggtttttaa tttgataaga ttttaatttg tttaagtttt 8100
ttaaataagg gttttaaatg ttttttagtcg ttttttttatt gaatttttttt taatttttttt 8160
aagattataa agtatatgtg taaagtaaat attttttttt attgtattgt tagtcgatga 8220

tttataatta agttaataag aatttagttt ttttttgttg aatgtgttta ttaattatat 8280
tttagttttt ttttttaaat tttagaatag ttgtggtttt tataatatta tgttttttaa 8340
agttttattt tatgaaggga ttttattata ttaaagaatg aaaaaaattt ttattgtagt 8400
tagtatatat agtttttttat tttttgtttt ttaagattta aattttagag ttgtaaatat 8460
ttttggaagt ttgggtgtta atgttttatt ttagaaagtc gagaagtttt atagagttat 8520
atagattttt aaatttattt tttataaatt tatagaattt tgataaaagt tttggtggtt 8580
ttattttatc gatggaattt ttattacgat aaatatatat gtatgaagga ttttaattag 8640
tttttaaagt ggttgaaaaa tttaagggta cgtgattgtt ttttatagtg ttaacgtgtg 8700
cgagatgttg gaagtattgg ggattagtag tagtttagat gtttaaaaag ataaggtgtt 8760
ttaatttgtg tggatttatt gaagttaagt ggtgaataaa gataattatt tagataattt 8820
agattaaagt aaaagtaaaa ttatatttat ttgtatatat atatttatat ttatttttata 8880
ttatagatat atatacgtat atatatattg gttttgtaaa taattgattt aaagtgagga 8940
ttttttttgt atttttttag taggagtttt aatatttttt taatttttta attatttat 9000
a 9001

<210> 3
<211> 9001
<212> DNA
<213> Artificial Sequence

<220>
<223> chemically treated genomic DNA (Homo sapiens)

<400> 3

tgtaaagtga ttaggagatt aggagaatgt tgaaattttt gttggaaaaa tgtaaagaaa 60
atttttattt tgagttagtt gtttatagag ttagtgtgtg tgtgcgtgtg tgtgtttgta 120
atataaaatg gatgtgaata tatatatata aatagatatg gttttgtttt tattttaatt 180
tgaattattt agataattgt ttttatttat tatttgattt taatgggttt atataaatta 240
ggatatttta ttttttttagg tatttaggtt gttgttgatt tttagtgttt ttaatatttc 300
gtatacgttg gtattatgag gagtagttac gtgtttttgg gttttttaat tattttggag 360
gttgattgag gtttttttata tatgtatatt tgtcgtgatg aaagttttat cggtagagtg 420
gagttattag agtttttatt aaaatttttgt gggtttatga gagatgggtt tagaaattta 480
tatggttttg tggggttttt cggtttttta aaataaggta ttaatattta agtttttaaa 540
aatatttgta gttttggggt ttgaattttg aaaaataagg agtgaggggt tgtgtatatt 600
aattatagtg gagatttttt ttatttttta atgtgatgga gttttttttat gaaatgaagt 660
tttaaggggt atggtattgt ggggattata gttattttga ggtttaaaag aagaaattgg 720
aatatgatta gtaaatatat ttagtagaaa agagttggat ttttattgat ttagttatag 780
gttatcggtt ggtagtgtaa tgggaggaaa tatttatttt atatatatat tttatgattt 840
tgggggaatt agaggaaatt taataagaaa acggttagaa atatttaaaa tttttattta 900
aaagatttaa gtaaattaga gttttattag attaaaaatt attataaatg taagagtatt 960
gtttttagtg aaacgttgtg gggtttgaga aggagatttt tcgttaaatt ttcgggataa 1020
aatgcgttat ttaagtatta gataatgagt agaatgtaaa ttaatttaat ttttttttatt 1080
aataggttgt tagtgtaatg tgtataattt agtgataaga ttgtaggatt taatatagtt 1140
ggatgtatga gttttagtta atgtagattt gttatatgag gatgtgtttt attttgagta 1200
ggtgtttgta tgtgtggaat ggggtaaagt ggaataaaag gttaaaagta gaaatgttga 1260
tttaaagttt attatgaaga aatttttttt ttgtagttaa attatttta aagtgggatg 1320
atattggtga agaaagattg aaaaataatt tttatgtgcg tttttggatt gtaagtttaa 1380
aatggggagg agttgtagat agggtttggg ggtggttagg gtaaaggaga gatatataag 1440
ttgtaaatat atttgtagtt tgttttattt attttgtttt atatcgaata agtttttttaa 1500
ttttgtgaat aaggataagg agggagtgtt ttaaagatat tttatgttgg tattgtaaat 1560
tattgattgt aatgttaaat aaatatatat ttagagatga taatattaat tttatagtaa 1620
aataatcgtt tatgtagaaa tttagaggag attagtttgt ttttttttagt tgatttatgt 1680
tgggggataa aaggattttt aaaaattatt ttgaatatgt ttggattttt ttttttaatt 1740
tttttggaaa ttaaatttgt ttggaaatag tgttataaag agttgatgtt tttaaaggtg 1800
atttttttttg ttttatataa ataaggtttt gtttttgtta gttgagcgta gttttaggtt 1860
tttcgttttt agtttatata tatttttttt gtttgtttgg attttaatgg tttaagatag 1920
ttttgagttt attgggaaaa gaaaatgatt gttaaaaatt atttttgaaa ttggttattt 1980
ggtaatattt ttaattgtat ggaaatttat taaggtatat tttatatata attagtttaa 2040
ggttgttgat tttataggtt ttatggattt aaatttgatt gataataaag taaataagag 2100
agtcgaattt aaagcgtggt tttttcgggt taggacgagt ttaatatagt gtataaggaa 2160
tttgaaagat ttaggatatg tgttttaatt aacgttaagt agaatggata agtttttagt 2220
attttgaaaa cgttgggtta gggttttttt tttattgtgt gtttttttgtt tggggattaa 2280

taagtattat agagaacgtg atttgaggcg atttttttatt tttgtataaa tttagagtga 2340
attattaaat agttgttcgt ttaaagttaa ggtaattttt ttttgacggg tttatttgtt 2400
tttcgatttt taatttatta gtttgttttt ttagggtttt gttttttttg taattaaagt 2460
ttttttagat tagcgtagta tttatttgat aggttgtttg gaaaatttaa gatcggagag 2520
gtgatttgtt gttgtttttt aaatttttta gttttaagta acgtgttttt ttttttatatg 2580
gggtggggga ttggaaatgg atgtagtgag atataaagag tgggtgtttt gttgattttt 2640
gtatttttttt ttttttgatt attttatttt tttttttttaa gttttcgatt tttagtttta 2700
ttttttttatt tttgggttcg tattaaaagt cggatcgttt tgggttgggt aggagttgaa 2760
ttttcgggag tttgtttgtg tagatttagt gcgtacggcg aggtagtagt tcggtttcgt 2820
attgttgata ggtgtaggta ggatagtttt tttatcgcgg ttcggggcgt tttgattggt 2880
gcggagttac gttagtcgta ttcggagaag ggtttgggag gaggcggagg cggagagggt 2940
tggggagggt cgcggcggag tgacgtttcg gtattaggaa gttcgttttt ggttttaaga 3000
tgttaggtta atagggaagc gcggagtcgt agatttggtt cgtcgttcgt ttgggtgttt 3060
ggagttgagt tgcggtaagg ttcggttttt gttcgatcgt tcgaggggtg tgcgtgtgcg 3120
cgttgcggag ggtgcgttta gagggtcgcg tcgtggttgt agcggttgtt gtcgtcgtag 3180
gggatttaat attatttatt tgtttttgtt atttttgata ttttttttgtt agggttgtcg 3240
cgtgggggggg gggcgggtag agcgcggtcg gcgttagttt tttttattgg aggggttttt 3300
gggggaggga gggagagaag aaggggttt ttgtttattt ttgtttcgtt ttggagtttg 3360
gaagtttgtt ttttaaagac gttttgagtg gtgtttttttt gtttatattt tatgttttcg 3420
tttgttcgtt gatttttcgt tttcggattt tttcgtttga gtttttcgga ggagacgggg 3480
gtagtttggt ttgagaattc ggcgggggtt gcgtttttttg gtttttttcg tagcggggaa 3540
atttcgcgtt tagagcgcga ttcggagcgg gtagcggcgg ttacgggggt tcggcggggt 3600
agtagttaag gattagtaga gcgtcgcgtt ttttcgttta tgaattgtat gaaaggttcg 3660
ttttatttgg agtatcgagt agcggggatt aagttgtcgg tcgtttttttt attttttttgt 3720
tattattttt agtcgttagt tatggtttcg gttttggttt tcggttagtt tcggtcgttg 3780
gatttttttta agtataggtt ggaggtgtat attattttcg atatttttag ttcggaggtc 3840
gtaggtaagg cgtcgcgtcg ttttgtagat attttcgttt agttgttttg cgttattcgt 3900
ttttttttcgt tttaaggaag ttagtttttt cgggggggagg cgtggtggga gtggtcgttc 3960
gtttggtttt tcgtagaatt ttcgggagtc ggaattttga ttatttcgta ttttttttagt 4020
ttttttttcga tcggttcggt ttttgggggcg ttaagggcgc gagtaatttt gtcgtttttt 4080
ttattcgtat tttggttttt ttttttgtttt ttgggttata aaaatttttag tattttgatt 4140
cgaggatttt tagaggtcgt cgatttttgt ttttgttttt ttttcggttt ttagttttcg 4200
aggagtttta ttcgttagga aattgtttga aattatttag aaatgttttt cgcgaagagg 4260
tatttttttt ttttttttgg gaaagggtcg gcgaatttcg gtgtttaatc gaatttttat 4320
attttttttt agtttttttta aatcgtatgg aaatttgagt tttttgcgag ggggaggggg 4380
gtttgtaaat tacgcgcgtg tgcgcgtttt aggagatttg gtgtgtttgc gtagaggtgt 4440
ataaatatat ttgaaagtat aggttataaa agtgaatgtg tcgttgtagt gagataaata 4500
tgtaaataaa acgtgcggcg ttgggggagg ggaggaaatg gggcgcggat atttatattt 4560
gcgtttgtat attttatagg cgtagcgttt ttcgcggttc ggagtcgtcg cgcgtatttt 4620
ttttcggcgt taggtagttt agtttttttta cggtttttgt cgtcggttta gttggcgttc 4680
gcgttgtagg tgggtatgtt gacgggaaag tgtgtgtgtt tcgtttttag agaaagataa 4740
aagttagtag gggaagaatg aggacgtggg cgtcgaggat tcgtttaaga agaagcggta 4800
aaggcggtag cggatttatt ttattagtta gtagttttag gagttggagg ttatttttta 4860
gaggaatcgt tattcggata tgtttatacg cgaagaaatc gttgtgtgga ttaattttac 4920
ggaagttcga gttcgggtag gagttagtac ggagtttggg agggatgggg ggaggatgtt 4980
gtggaggtat aggttaagta gattaggaga gaatgtggaa ggtagcgtcg tttgggaggg 5040
cgtcggtggg gcgtagtttt gtaaaggtag aaggtttcgc ggcggtttgg ttgcgagatt 5100
atagtttttt tttcgaggtc gataggattg tcgttttggt ttaggttttt agagcggtat 5160
cggtttattg tttcgttatt tcgcgatttt acgagttggg ttgtatgggt aatttttttgt 5220
ataggatatt gtgtttttgg tttgtagttg ttagagtaga gttaataaaa tttttattag 5280
gttaagagtc gcgaataggt tttaatttgt gagtttttaa taaggaaaat tcgttagaga 5340
tacggaagag ttggtttttt ttgggaaatt tttgtttcgg ttttggttta gtttttttttt 5400
ttttgggttc gcgtttttta tatttttttt acggttgttt cggttattta ggtttttttt 5460
atatatttta ttttttagtt ttgtgatttt cgggagtaaa gttttaatat ataattatta 5520
gtttttttag aaggagaaag aaaaaaaagaa gaaagatttt tttgtttggt ttatttattt 5580
ttttttagga gttgaatttt ggaaattgaa atttatattt ttttttttaa attataatta 5640
tagttttgta aaaagggttt attttaattt tgtagtaaat ttgtatttta tggattggta 5700
aaaatgagtt taaataaata atttaatagt aacgttttgg tttatgttgg tcggtggaag 5760
attttaaatt tgttaggatt ttggaagtag aaaatagaat taagtaaatt aagcggtatt 5820
tagaggtttt gttgttaaaa aaaaaaaatt aagtgttttg ggtagaaaaa ataaagtttt 5880
cggttagagt agagtaaata aaaagaagaa aataacgata aaaagaataa agattaaaat 5940
gtttttttaa attagaggga atgaagatat tttttgggtg gtatttgtgt aaggtatgag 6000
gttatgttgg tggataaaag gtcgggaaga agttgaaaat ggttttagtt taattgttta 6060

gagttagagt tgggttttgg gcggcgtggt tttgagtaag gttagttttt tattagtttt 6120
tttgtatatt aagggaacgg gtttttttacg tatttttttc gtttgagtaa agtttagatg 6180
gtttagggta gaaatggtaa gtaattaaag atagagttta tgggtttttt gggatttttc 6240
gaaaacgttt ttttatttcg ttcgttattt cgtagtttta ttttagtgtt ttgtagtcgc 6300
ggcgttgggt ttttttttgta gttgtttttt tttttagggc ggttgtttgt cgagttaagt 6360
gggagtgagg cgtgtttttt atagtagtcg ggtgtaaaga ggaaggggga taaaaaggaa 6420
attaagaatg aaaggaaaaa gagaaaaagc ggattatacg gttgggttcg gcggagatgt 6480
gtaatgtgaa atattattgg tgttagttcg gatattttag gttaggtttt tttttaatat 6540
ataaaagtcg tcgtttgggg cgatagggag gttcgatgtg gattgggatc ggggttgcgg 6600
ttgggttatc ggatacgggt ggaagtcggt cggtttgggt ggtcgtttgt aaagttaaac 6660
gattcggttg ggtttggcgc gcggataggt ttgtggtggg tttagggtaa agaagaggta 6720
gagcgaaaga agggggaatt tttaaaatta ttttttttcgg gttttcggag tttaatatgt 6780
taagtttttg gagttaacga gttgacgaag aggtggtttt ttgtttttta tttggttgtt 6840
ttgttaggcg agaaagagtg ttggcggttt agtttttgtt aagggagtac gtattagggg 6900
gtgggggacg atagtggagg ttagggaagg aagggaggaa ttgcgtggga gaaagagcga 6960
ttttttagtg ttttttttagt ttttttttttt tattcgtggg tttgtggttt tggaatggaa 7020
gtaagtttgt aaggtgtttc gggaagggtt ggaaaagttt gttgtttcgc gtttgtttta 7080
tattaagtgt ttttggattt ggagaaacgt ttggttgagt gattaaatcg ttcgtaggtt 7140
tttatgcgtt cggttgaggt ttgtggcgta gtttcgagtt ttagttcgta ggttagagta 7200
gattaggttt tttgcgtttg gtggagattc gggttagtaa ttgaaagttg gttttggtat 7260
tttggtgtgt agggcggtgt agtgaagcga ggttagggtg tgtgagtgcg ttagcgtgtg 7320
tgtcggggga aggcgggggt tggttttcga tggaagtttt agtaatttgt attgtggtat 7380
ttgtttgttt ttttgttttta atcgtttttta ggtttggttt aagaatcgtc gggttaaatg 7440
gagaaagagg gagcgtaatt agtaggtcga gttatgtaag aatggtttcg ggtcgtagtt 7500
taatgggttt atgtagtttt acgacgatat gtatttaggt tatttttata ataattgggt 7560
cgttaagggt tttatattcg ttttttttatt tattaagagt ttttttttttt ttaattttat 7620
gaacgttaat tttttgttat tatagagtat gtttttttta tttaatttta tttcgtttat 7680
gagtatgtcg tttagtatgg tgtttttagt agtgataggc gtttcgggtt ttagttttaa 7740
tagtttgaat aatttgaata atttgagtag ttcgtcgttg aatttcgcgg tgtcgacgtt 7800
tgtttgtttt tacgcgtcgt cgatttttttc gtatgtttat agggatacgt gtaattcgag 7860
tttggttagt ttgagattga aagtaaagta gtattttagt ttcggttacg ttagcgtgta 7920
gaattcggtt tttaatttga gtgtttgtta gtatgtagtg gatcggttcg tgtgagtcgt 7980
atttatagcg tcgggatttt aggatttttgt cggatggggt aatttcgttt ttgaaagatt 8040
gggaattatg ttagaaggtc gtgggtatta aagaaaggga gagaaagaga agttatatag 8100
agaaaaggaa attattgaat taaagagaga gtttttttga ttttaaaggg atgtttttag 8160
tgtttgatat tttttattat aagtatttttt aatagttgta aggatatata tataaataaa 8220
tgtttgattg gatatgatat tttaatatta ttataagttt gttatttttt aagtttagta 8280
ttgttaatat ttaaatgatt gaaaggatgt atatatatcg aaatgttaaa ttaattttat 8340
aaaagtagtt gttagtaata ttataatagt gtttttaaag gttaggtttt aaaataaagt 8400
atgttatata gaagcgatta ggatttttcg tttgcgagta agggagtgta tatattaaat 8460
gttatattgt atgtttttaa tatattatta ttattataaa aaatgtgtga atattagttt 8520
tagaatagtt tttttggtgg atgtaatgat gtttttgaaa ttgttatgta taatttattt 8580
tgtgtataat atttcgtata atattattgt tttatttttt agtaaatatg aaataaatgt 8640
gttttatttt atgggagtaa aatatattgt atataaattg gtttggattt ttttttttttt 8700
ttttttgttat taatttggtt aggatatttt agttattgtt ttttaaataa attagttttt 8760
tttgtttgtt tagttaaata tataaggtag tagtttttat ttaaatttgg tagaaataaa 8820
tgatagttat ttattagaaa ttaaaaaagaa aaaaaaaggt attttcgggg gggaaaaggg 8880
ttataaaatt taattttgtt tttttaattt ttttttggtt taaatttaga ggattttatt 8940
atggttagta aataatatga aaaagaaaaa agaagaaaga aatttagtaa gtttattagt 9000
t 9001

<210> 4
<211> 9001
<212> DNA
<213> Artificial Sequence

<220>
<223> chemically treated genomic DNA (Homo sapiens)

<400> 4

agttgatgga tttgttaaat ttttttttttt ttttttttttt tttatattat ttgttagtta 60
taatggaatt ttttaggttt aagttaaaga aaaattggag agataaaatt agattttgta 120

gttttttttt ttttgggaa tgttttttt ttttttttta gttttgatg aatggttatt 180
atttatttt attaaattta aataaggatt gttgttttgt atgtttaatt aggtaggtag 240
agggaattgg tttgtttagg aagtagtgat tgagatgttt tggttaagtt agtgatagag 300
gaggggagaa agaatttaga ttaatttgta tgtagtatat tttatttta tgaaataaaa 360
tatatttgtt ttatatttgt tgaaaagtaa aataataata ttgtatgaaa tgttatatat 420
agggtaggtt gtatatagta gttttagaaa tattattgta tttattagag aaattatttt 480
aaaattgata tttatatatt ttttataata ataataatat gttagaaata tatagtgtgg 540
tatttagtat atatattttt ttgtttgtaa gtgaaaaatt ttaattgttt ttgtataata 600
tgttttattt taaagtttaa tttttaaaaa tattgttgtg atattattaa taattgtttt 660
tataaaatta atttgatatt ttgatatata tatatttttt tagttattta aatgttaata 720
atgttaaatt taaaaaataa taagtttata gtaatgttaa aatgttatat ttagttaaat 780
atttgtttgt gtatgtgttt ttgtaattgt tagaaatatt tgtagtgaaa gatgttagat 840
attgaggata ttttttgaa attaaaggag ttttttttt gatttagtgg ttttttttt 900
tttatatagt ttttttttt ttttttttt ttagtgttta tgattttta gtataatttt 960
tagtttttta agggtggagt tgttttattt ggtaaggttt taggattttg gtgttgtggg 1020
tgtggtttat atgggttggt ttattgtata ttggtaagta tttaggttgg aggttgggtt 1080
ttgtatgttg gtgtagttga agttggagtg ttgttttgtt tttagtttta ggttggttag 1140
gtttgagtta tatgtgtttt tataaatata tggaggagtt ggtggtgtgt aaggataggt 1200
aggtgttggt attgtggaat ttagtgatgg gttatttagg ttgtttaagt tatttaggtt 1260
gttgagattg gagtttggga tgtttgttat tgttgagggt attatgttgg atgatatgtt 1320
tatggatgag atagagttgg gtggggaaaa tatgttttgt gatgataggg ggttgatgtt 1380
tatagagttg aagaagggga agtttttggt ggatagggag gtggatgtaa ggtttttggt 1440
ggtttagttg ttgtaggaat agtttgggta tatgttgttg tagggttgta tgagtttatt 1500
gaattgtggt ttgaagttat ttttgtatag tttggtttgt tggttgtgtt ttttttttt 1560
ttatttggtt tgatgatttt tgaattaaat ttgggggtgg ttggggtaag ggagtaaata 1620
gatgttatag tgtagattat taaaattttt attggaggtt aatttttgtt ttttttgat 1680
atatatgtta gtgtatttat atattttggt tttgttttat tgtattgttt tgtatattaa 1740
gatattaggg ttagttttta gttattggtt tgggttttta ttaagtgtag gagatttggt 1800
ttgttttggt ttgtgagttg ggatttggag ttatgttata aattttagtt gaatgtatgg 1860
agatttgtgg atggtttgat tatttagtta ggtgtttttt taggtttaaa aatatttaat 1920
gtaaaataaa tgtggggtag taggtttttt taatttttt tggggtattt tgtaaatttg 1980
tttttatttt aaagttatag atttatggat gaggagaagg ggttggaagg gtattagagg 2040
attgtttttt tttttatgta attttttttt ttttttttg atttttattg ttgttttta 2100
ttttttggta tgtgtttttt taatagggat taggttgtta atatttttt ttgtttagta 2160
aaataattaa ataaagagta aaagattatt tttttgttag tttgttaatt ttaggagttt 2220
ggtatattaa attttgggaa tttggaaagg gtagttttgg agatttttt ttttttgtt 2280
ttgtttttt tttattttaa gtttattata ggtttgtttg tgtgttaggt ttagttgggt 2340
tgtttggttt tgtaggtggt tatttaggtt ggttggtttt tatttgtgtt tggtggttta 2400
gttgtaattt tgattttaat ttatattggg ttttttttgtt gttttagatg gtggttttg 2460
tgtattggag agaggtttgg tttgagatat ttgagttgat attagtgatg ttttatatta 2520
tatatttttg ttgggtttag ttgtgtaatt tgtttttttt ttttttttt ttatttttga 2580
ttttttttt attttttttt ttttttgtat ttgattgtta taaaaagtat gttttatttt 2640
tatttggttt gataagtagt tgttttggaa ggagaggtag ttgtaaggag agtttagtgt 2700
tgtggttata aagtattagg gtggagttgt ggaatagtgg gtggggtggg agggtgtttt 2760
tgaaggattt tagaaaattt atagattttg tttttaatta tttgttattt ttattttagg 2820
ttatttaaat tttgtttagg tgagaagagt atgtgagagg tttgtttttt tgatgtgtaa 2880
gagagttaat gaaagattga ttttgtttaa aattatgttg tttaggattt agttttggtt 2940
ttggatagtt aaattaaaat tatttttaat ttttttttgg ttttttattt attagtatag 3000
ttttatgttt tgtataaatg ttatttagag agtgttttta ttttttttga tttgggagag 3060
tattttggtt tttatttttt ttattgttgt ttttttttttt ttgtttgttt tgttttaatt 3120
gggggttttta ttttttttat ttagagtatt taatttttttt tttttaatag taaagttttt 3180
ggatgttgtt tgatttgttt gattttgttt tttgttttta gaattttaat aaatttggaa 3240
ttttttattg attagtataa attaggatgt tgttattggg ttatttattt gagtttattt 3300
ttgttaattt ataaagtata gatttgttat aaagttaagg taagtttttt ttataaaatt 3360
atgattataa tttagaagag ggggtgtgag ttttaatttt tagagtttaa tttttgagag 3420
aagataaata aattaagtag aaaagttttt ttttttttttt ttttttttttt tttaagagga 3480
ttagtagttg tgtattaaaa ttttgttttt ggagattata aaattaggaa atagggtgtg 3540
tgggagagat ttgaatggtt gaaataattg taaagaaggt gtaagaagtg tgagtttagg 3600
agggaaaaag ttgggttagg gttgggataa aggtttttta gggagggtta attttttttgt 3660
gtttttggtg ggtttttttt gttaaaggtt tataggttgg agtttgtttg tggtttttgg 3720
tttggtaggg attttattag ttttgttttg gtaattgtaa gttaggaata taatgttttg 3780
tgtaggggat tgtttatgta gtttagtttg tgagattgtg ggatggtggg gtagtgagtt 3840
ggtgttgttt tgggagtttg agttagggtg gtagttttgt tggttttgga gagggaattg 3900

taattttgta attaggttgt tgtgaggttt tttgtttttg taaagttgtg ttttattggt 3960
gttttttag gtggtgttgt tttttatatt ttttttggt ttatttggtt tgtattttta 4020
taatattttt tttttatttt ttttagattt tgtgttggtt tttatttgga tttgggtttt 4080
tgtaaggttg gtttatatag tgatttttt gtgtgtggat atgtttgggt agtggttttt 4140
ttggaaagtg gtttttagtt tttggagttg ttggttggta aagtgagttt gttgttgttt 4200
ttgttgtttt tttttagatg ggtttttggt gtttatgttt ttatttttt tttgttggtt 4260
tttatttttt tttgaaaatg aaatatatat attttttgt tagtatgttt atttgtaatg 4320
tggatgttaa ttggattggt ggtagaagtt gtggaagagt tgggttgttt ggtgttggag 4380
gagggtgtgt gtggtggttt tgggttgtga ggagtgttgt gtttgtgggg tgtgtaggtg 4440
taagtgtggg tgtttgtgtt ttatttttt ttttttttta gtgttgtatg ttttatttat 4500
atgtttattt tattgtagtg gtatatttat ttttatagtt tgtgttttta agtatattta 4560
tatatttttg tgtagatata ttaaattttt tgggatgtgt atatgtgtgt ggtttataga 4620
ttttttttt ttttgtagaa agtttagatt tttatgtggt ttgggaaggt taggaaaaga 4680
tgtggggatt tggttgggta ttgaagtttg ttggtttttt tttaaaaaaa aaaaaaaaat 4740
gtttttttgt gaagggtatt tttgagtggt tttaggtaat tttttaatga gtggagtttt 4800
ttgggagttg aaagttgaga ggaaaatagg gatagaggtt ggtggttttt gaaggttttt 4860
gaattaagat gttgggattt ttgtgattta ggaaatagaa gggaggttag ggtatgaata 4920
gagagggtgg tagaattgtt tgtgtttttta gtgttttagg agttgggttg gttgagggag 4980
aattaaaggg atgtggggta gttaaaattt tggtttttgg aagttttgtg gggagttagg 5040
tgaatgatta ttttttattat gttttttttt ggaggggttg attttttttgg ggtgagaggg 5100
agtgggtggt gtagagtagt tgagtgggaa tgtttgtagg gtggtgtggt gttttatttg 5160
tggtttttgg gttggaggtg ttggagatgg tgtgtatttt tagtttgtgt ttggaggagt 5220
ttagtgattg gggttgattg ggagttagaa ttgaagttat ggttaatggt tggggatggt 5280
gataggaaga tgaggagatg gttgatagtt tggtttttgt tgtttggtgt tttaagtgaa 5340
gtgggttttt tatgtagttt atggatgagg gagtgtgatg ttttattagt ttttggttat 5400
tgttttgttg agtttttgta gttgttgttg tttgttttgg gttgtgtttt aggtgtggag 5460
ttttttttgtt gtggggagag ttaggggatg taatttttgt tgagttttta agttaagttg 5520
tttttgtttt ttttggaagg tttaagtgaa aaagtttgga gatggaaagt tagtgggtaa 5580
atgaagatat gggatgtggg tagaagggta ttatttagag tgtttttagg gagtaggttt 5640
ttaagtttta aagtgaaata agagtgggta aagatttttt ttttttttttt ttttttttttt 5700
taagaatttt tttaataagg aaagttaatg ttgattgtgt tttgtttgtt ttttttttat 5760
gtggtagttt tgatagagaa gtgttaagag tgatagggat aggtaggtga tattagattt 5820
tttgtggtgg tagtagttgt tgtagttatg atgtggtttt ttgagtgtat tttttgtaat 5880
gtgtatatgt atatttttttg ggtggttgaa taggagttgg gttttgttgt agtttagttt 5940
taggtattta ggtgagtgat ggattagatt tgtggttttg tgttttttttg ttggtttaat 6000
attttaaaat tagaggtggg ttttttggtg ttgagatgtt attttgttgt ggtttttttt 6060
agtttttttt gtttttgttt ttttttagat ttttttttgg gtgtgattga tgtggttttg 6120
tattaattag gatgttttga gttgtggtgg agggattgtt ttgtttgtat ttattagtag 6180
tgtggggttg ggttattgtt ttgttgtgtg tattgggttt atataggtaa gtttttggga 6240
atttagtttt tgtttagttt aaggtgattt ggtttttagt atgaatttaa aggtgaagag 6300
atgaggttag gagttgaagg tttgggagaa gagagtggaa tggttaagaa gagaaaggta 6360
taaggattaa taagatattt attttttgtg ttttattata tttatttta attttttatt 6420
ttatataaaa aggagatatg ttatttaaaa ttagaaaatt tgaaaaatag taataaatta 6480
ttttttttgat tttaaatttt ttaaatagtt tgttaagtga atgttgtgtt aatttgaaga 6540
agttttaatt gtaaagaaga tagagttttg aaaaggtagg ttaataaatt agaaattgag 6600
aagtaaatgg atttgttaaa agaaaattat tttgatttta aatgaataat tgtttggtgg 6660
tttatttttgg atttatataa gaataaaaag ttgttttaga ttatgttttt tgtgatgttt 6720
attagttttt agatagaaaa tatataatag aagagaaatt ttaatttagt gtttttaaaa 6780
tgttgaaagt ttatttattt tatttaatgt tgattaagat atatatttta gatttttttaa 6840
atttttttgta tattgtatta agtttgtttt aatttgagag agttatgttt taaatttgat 6900
ttttttgttt attttattat taattagatt taaatttata aagtttgtag aattaataat 6960
tttgagttaa ttatatatga aatatgtttt aatgaatttt tatataatta agaatgttgt 7020
taaataatta attttaagga taatttttaa tagttatttt ttttttttag tgagtttaag 7080
gttgttttga gttattaaag tttaagtagg tagaaggggt gtgtgtgagt taagggtgaa 7140
aagtttagaa ttgtgtttaa ttagtaaaag taaaatttta tttatataaa ataaaaaaaa 7200
ttatttttgg agatattaat tttttatagt attgttttta agtaaattta atttttaaag 7260
aaattaaaga aagaaattta aatatattta aaataatttt tgaaagtttt tttgtttttt 7320
agtataggtt agttggagag gataaattaa ttttttttgg gtttttgtat gggtgattgt 7380
tttattatgg agttagtgtt attatttttg aatgtgtatt tgtttgatat tatagttaat 7440
gatttgtaat gttagtatga agtatttttа aaatattttt tttttgtttt tgtttataag 7500
attgggaaat ttatttgatg tggaataaag tggatgaagt agattataaa tatatttgta 7560
atttatgtgt ttttttttg ttttgattat ttttaaattt tatttgtaat ttttttttat 7620
tttaaatttg tagtttaaag atgtatatga gaattgtttt ttagtttttt tttattagta 7680

ttatttttatt ttaagaataa tttagttgta agggaggaat ttttttatag taagttttaa 7740
attagtattt ttgtttttaa ttttttattt tattttattt tattttatat atatagatat 7800
ttgtttagag taaaatatat ttttatgtga taggtttgta ttagttgagg tttatatatt 7860
tagttatatt aggttttgta attttattat taaattatat atattatatt agtagtttgt 7920
tggtaaagaa ggttaaatta atttatattt tgtttattat ttggtgttta aatgatgtat 7980
tttatttttgg agatttggtg gagaattttt tttttagatt ttatagtgtt ttattgaaga 8040
taatgttttt atatttgtag tggtttttaa tttgataaga ttttaatttg tttaagtttt 8100
ttaaataagg gttttaaatg tttttagttg ttttttttatt gaatttttttt taatttttttt 8160
aagattataa agtatatgtg taaagtaaat attttttttt attgtattgt tagttgatga 8220
tttataatta agttaataag aatttagttt ttttttgttg aatgtgttta ttaattatat 8280
tttagttttt ttttttaaat tttagaatag ttgtggtttt tataatatta tgttttttaa 8340
agttttattt tatgaaggga ttttattata ttaaagaatg aaaaaaattt ttattgtagt 8400
tagtatatat agtttttttat tttttgtttt ttaagattta aattttagag ttgtaaatat 8460
ttttggaagt ttgggtgtta atgttttatt ttagaaagtt gagaagtttt atagagttat 8520
atagattttt aaatttattt tttataaatt tatagaattt tgataaaagt tttggtggtt 8580
ttatttttatt gatggaattt ttattatgat aaatatatat gtatgaagga ttttaattag 8640
tttttaaagt ggttgaaaaa tttaagggta tgtgattgtt ttttatagtg ttaatgtgtg 8700
tgagatgttg gaagtattgg ggattagtag tagtttagat gtttaaaaag ataaggtgtt 8760
ttaatttgtg tggatttatt gaagttaagt ggtgaataaa gataattatt tagataattt 8820
agattaaagt aaaagtaaaa ttatatttat ttgtatatat atatttatat ttatttttata 8880
ttatagatat atatatgtat atatatattg gttttgtaaa taattgattt aaagtgagga 8940
ttttttttgt atttttttag taggagtttt aatatttttt taatttttta attattttat 9000
a 9001

<210> 5
<211> 9001
<212> DNA
<213> Artificial Sequence

<220>
<223> chemically treated genomic DNA (Homo sapiens)

<400> 5

tgtaaagtga ttaggagatt aggagaatgt tgaaattttt gttggaaaaa tgtaaagaaa 60
atttttattt tgagttagtt gtttatagag ttagtgtgtg tgtgtgtgtg tgtgtttgta 120
atataaaatg gatgtgaata tatatatata aatagatatg gttttgtttt tattttaatt 180
tgaattattt agataattgt ttttatttat tatttgattt taatgggttt atataaatta 240
ggatatttta ttttttttagg tatttaggtt gttgttgatt tttagtgttt ttaatatttt 300
gtatatgttg gtattatgag gagtagttat gtgtttttgg gttttttaat tattttggag 360
gttgattgag gttttttata tatgtatatt tgttgtgatg aaagttttat tggtagagtg 420
gagttattag agtttttatt aaaatttttgt gggtttatga gagatgggtt tagaaattta 480
tatggttttg tggggttttt tggtttttta aaataaggta ttaatattta agtttttaaa 540
aatatttgta gttttggggt ttgaattttg aaaaataagg agtgaggggt tgtgtatatt 600
aattatagtg gagatttttt ttatttttta atgtgatgga gttttttttat gaaatgaagt 660
tttaaggggt atggtattgt ggggattata gttattttga ggtttaaaag aagaaattgg 720
aatatgatta gtaaatatat ttagtagaaa agagttggat ttttattgat ttagttatag 780
gttattggtt ggtagtgtaa tgggaggaaa tatttatttt atatatatat tttatgattt 840
tgggggaatt agaggaaatt taataagaaa atggttagaa atatttaaaa tttttattta 900
aaagatttaa gtaaattaga gttttattag attaaaaatt attataaatg taagagtatt 960
gtttttagtg aaatgttgtg gggtttgaga aggagatttt ttgttaaatt tttgggataa 1020
aatgtgttat ttaagtatta gataatgagt agaatgtaaa ttaatttaat ttttttttatt 1080
aataggttgt tagtgtaatg tgtataattt agtgataaga ttgtaggatt taatatagtt 1140
ggatgtatga gttttagtta atgtagattt gttatatgag gatgtgtttt attttgagta 1200
ggtgtttgta tgtgtggaat ggggtaaagt ggaataaaag gttaaaagta gaaatgttga 1260
tttaaagttt attatgaaga aatttttttt ttgtagttaa attatttta aagtgggatg 1320
atattggtga agaaagattg aaaaataatt tttatgtgtg tttttggatt gtaagtttaa 1380
aatggggagg agttgtagat agggtttggg ggtggttagg gtaaaggaga gatatataag 1440
ttgtaaatat atttgtagtt tgttttattt attttgtttt atattgaata agttttttaa 1500
ttttgtgaat aaggataagg agggagtgtt ttaaagatat tttatgttgg tattgtaaat 1560
tattgattgt aatgttaaat aaatatatat ttagagatga taatattaat tttatagtaa 1620
aataattgtt tatgtagaaa tttagaggag attagtttgt ttttttttagt tgatttatgt 1680
tgggggataa aaggattttt aaaaattatt ttgaatatgt ttggattttt ttttttaatt 1740

ttttttggaaa ttaaatttgt ttggaaatag tgttataaag agttgatgtt tttaaaggtg 1800
atttttttttg ttttatataa ataaggtttt gtttttgtta gttgagtgta gttttaggtt 1860
ttttgttttt agtttatata tatttttttt gtttgtttgg attttaatgg tttaagatag 1920
ttttgagttt attgggaaaa gaaaatgatt gttaaaaatt atttttgaaa ttggttattt 1980
ggtaatattt ttaattgtat ggaaatttat taaggtatat tttatatata attagtttaa 2040
ggttgttgat tttataggtt ttatggattt aaatttgatt gataataaag taaataagag 2100
agttgaattt aaagtgtggt ttttttgggt taggatgagt ttaatatagt gtataaggaa 2160
tttgaaagat ttaggatatg tgttttaatt aatgttaagt agaatggata agtttttagt 2220
attttgaaaa tgttgggtta gggttttttt tttattgtgt gttttttgtt tggggattaa 2280
taagtattat agagaatgtg atttgaggtg atttttttatt tttgtataaa tttagagtga 2340
attattaaat agttgtttgt ttaaagttaa ggtaattttt ttttgatggg tttatttgtt 2400
ttttgatttt taatttatta gtttgttttt ttagggtttt gttttttttg taattaaagt 2460
ttttttagat tagtgtagta tttatttgat aggttgtttg gaaaatttaa gattggagag 2520
gtgatttgtt gttgtttttt aaatttttta gttttaagta atgtgttttt ttttttatatg 2580
gggtgggggga ttggaaatgg atgtagtgag atataaagag tgggtgtttt gttgattttt 2640
gtatttttttt ttttttgatt attttatttt ttttttttaa gtttttgatt tttagtttta 2700
ttttttttatt tttgggtttg tattaaaagt tggattgttt tgggttgggt aggagttgaa 2760
tttttgggag tttgtttgtg tagatttagt gtgtatggtg aggtagtagt ttggttttgt 2820
attgttgata ggtgtaggta ggatagtttt tttattgtgg tttggggtgt tttgattggt 2880
gtggagttat gttagttgta tttggagaag ggtttgggag gaggtggagg tggagagggt 2940
tggggagggt tgtggtggag tgatgttttg gtattaggaa gtttgttttt ggttttaaga 3000
tgttaggtta atagggaagt gtggagttgt agatttggtt tgttgtttgt ttgggtgttt 3060
ggagttgagt tgtggtaagg tttggttttt gtttgattgt ttgaggggtg tgtgtgtgtg 3120
tgttgtggag ggtgtgttta gagggttgtg ttgtggttgt agtggttgtt gttgttgtag 3180
gggatttaat attatttatt tgtttttgtt atttttgata ttttttttgtt agggttgttg 3240
tgtgggggggg gggtgggtag agtgtggttg gtgttagttt tttttattgg aggggttttt 3300
ggggggaggga gggagagaag aaggggggttt ttgtttattt ttgttttgtt ttggagtttg 3360
gaagtttgtt ttttaaagat gttttgagtg gtgtttttttt gtttatattt tatgtttttg 3420
tttgtttgtt gatttttttgt ttttggattt ttttgtttga gtttttttgga ggagatgggg 3480
gtagtttggt ttgagaattt ggtgggggtt gtgtttttttg gttttttttg tagtggggaa 3540
attttgtgtt tagagtgtga tttggagtgg gtagtggtgg ttatgggggt ttggtggggt 3600
agtagttaag gattagtaga gtgttgtgtt tttttgttta tgaattgtat gaaaggtttg 3660
ttttatttgg agtattgagt agtggggatt aagttgttgg ttgtttttttt attttttttgt 3720
tattattttt agttgttagt tatggttttg gttttggttt ttggttagtt ttggttgttg 3780
gatttttttta agtataggtt ggaggtgtat attatttttg atatttttag tttggaggtt 3840
gtaggtaagg tgttgtgttg ttttgtagat attttttgttt agttgttttg tgttatttgt 3900
ttttttttgt tttaaggaag ttagtttttt tgggggggagg tgtggtggga gtggttgttt 3960
gtttggtttt ttgtagaatt tttgggagtt ggaattttga ttattttgta ttttttttagt 4020
ttttttttga ttggtttggt ttttgggggtg ttaagggtgt gagtaatttt gttgtttttt 4080
ttatttgtat tttggttttt ttttttgtttt ttgggttata aaaattttag tattttgatt 4140
tgaggatttt tagaggttgt tgatttttgt ttttgttttt ttttttggttt ttagtttttg 4200
aggagtttta tttgttagga aattgtttga aattatttag aaatgttttt tgtgaagagg 4260
tatttttttt ttttttttgg gaaagggttg gtgaattttg gtgtttaatt gaatttttat 4320
attttttttt agttttttta aattgtatgg aaatttgagt ttttttgtgag ggggagggggg 4380
gtttgtaaat tatgtgtgtg tgtgtgtttt aggagatttg gtgtgtttgt gtagaggtgt 4440
ataaatatat ttgaaagtat aggttataaa agtgaatgtg ttgttgtagt gagataaata 4500
tgtaaataaa atgtgtggtg ttgggggagg ggaggaaatg gggtgtggat atttatattt 4560
gtgtttgtat attttatagg tgtagtgttt tttgtggttt ggagttgttg tgtgtatttt 4620
tttttggtgt taggtagttt agtttttttta tggtttttgt tgttggttta gttggtgttt 4680
gtgttgtagg tgggtatgtt gatgggaaag tgtgtgtgtt ttgtttttag agaaagataa 4740
aagttagtag gggaagaatg aggatgtggg tgttgaggat ttgtttaaga agaagtggta 4800
aaggtggtag tggatttatt ttattagtta gtagttttag gagttggagg ttatttttta 4860
gaggaattgt tatttggata tgtttatatg tgaagaaatt gttgtgtgga ttaattttat 4920
ggaagtttga gtttgggtag gagttagtat ggagtttggg agggatgggg ggaggatgtt 4980
gtggaggtat aggttaagta gattaggaga gaatgtggaa ggtagtgttg tttgggaggg 5040
tgttggtggg gtgtagtttt gtaaaggtag aaggttttgt ggtggtttgg ttgtgagatt 5100
atagtttttt ttttgaggtt gataggattg ttgttttggt ttaggttttt agagtggtat 5160
tggtttattg ttttgttatt ttgtgatttt atgagttggg ttgtatgggt aatttttttgt 5220
ataggatatt gtgtttttgg tttgtagttg ttagagtaga gttaataaaa tttttattag 5280
gttaagagtt gtgaataggt tttaatttgt gagtttttaa taaggaaaat ttgttagaga 5340
tatggaagag ttggtttttt ttgggaaatt tttgttttgg ttttggttta gttttttttt 5400
ttttgggttt gtgtttttta tatttttttt atggttgttt tggttattta ggttttttttt 5460
atatatttta ttttttagtt ttgtgatttt tgggagtaaa gttttaatat ataattatta 5520

gtttttttag aaggagaaag aaaaaaagaa gaaagatttt tttgtttggt ttatttattt 5580
ttttttagga gttgaatttt ggaaattgaa atttatattt ttttttttaa attataatta 5640
tagttttgta aaaagggttt attttaattt tgtagtaaat ttgtatttta tggattggta 5700
aaaatgagtt taaataaata atttaatagt aatgttttgg tttatgttgg ttggtggaag 5760
attttaaatt tgttaggatt ttggaagtag aaaatagaat taagtaaatt aagtggtatt 5820
tagaggtttt gttgttaaaa aaaaaaaatt aagtgttttg ggtagaaaaa ataaagtttt 5880
tggttagagt agagtaaata aaaagaagaa aataatgata aaaagaataa agattaaaat 5940
gtttttttaa attagaggga atgaagatat tttttgggtg gtatttgtgt aaggtatgag 6000
gttatgttgg tggataaaag gttgggaaga agttgaaaat ggttttagtt taattgttta 6060
gagttagagt tgggttttgg gtggtgtggt tttgagtaag gttagttttt tattagtttt 6120
tttgtatatt aagggaatgg gtttttatg tatttttttt gtttgagtaa agtttagatg 6180
gtttagggta gaaatggtaa gtaattaaag atagagttta tgggtttttt gggatttttt 6240
gaaaatgttt ttttatttttg tttgttattt tgtagtttta ttttagtgtt ttgtagttgt 6300
ggtgttgggt tttttttgta gttgtttttt ttttttagggt ggttgtttgt tgagttaagt 6360
gggagtgagg tgtgtttttt atagtagttg ggtgtaaaga ggaaggggga taaaaaggaa 6420
attaagaatg aaaggaaaaa gagaaaaagt ggattatatg gttgggtttg gtggagatgt 6480
gtaatgtgaa atattattgg tgttagtttg gatattttag gttaggtttt tttttaatat 6540
ataaaagttg ttgtttgggg tgatagggag gtttgatgtg gattgggatt ggggttgtgg 6600
ttgggttatt ggatatgggt ggaagttggt tggtttgggt ggttgtttgt aaagttaaat 6660
gatttggttg ggtttggtgt gtggataggt ttgtggtggg tttagggtaa agaagaggta 6720
gagtgaaaga agggggaatt tttaaaatta ttttttttgg gtttttggag tttaatatgt 6780
taagtttttg gagttaatga gttgatgaag aggtggtttt ttgtttttta tttggttgtt 6840
ttgttaggtg agaaagagtg ttggtggttt agtttttgtt aagggagtat gtattagggg 6900
gtgggggatg atagtggagg ttagggaagg aagggaggaa ttgtgtggga gaaagagtga 6960
ttttttagtg ttttttttagt ttttttttttt tatttgtggg tttgtggttt tggaatggaa 7020
gtaagtttgt aaggtgtttt gggaagggtt ggaaaagttt gttgttttgt gtttgtttta 7080
tattaagtgt ttttggattt ggagaaatgt ttggttgagt gattaaattg tttgtaggtt 7140
tttatgtgtt tggttgaggt ttgtggtgta gttttgagtt ttagtttgta ggttagagta 7200
gattaggttt tttgtgtttg gtggagattt gggttagtaa ttgaaagttg gttttggtat 7260
tttggtgtgt agggtggtgt agtgaagtga ggttagggtg tgtgagtgtg ttagtgtgtg 7320
tgttggggga aggtgggggt tggtttttga tggaagtttt agtaatttgt attgtggtat 7380
ttgtttgttt ttttgttttta attgttttta ggtttggttt aagaattgtt gggttaaatg 7440
gagaaagagg gagtgtaatt agtaggttga gttatgtaag aatggttttg ggttgtagtt 7500
taatgggttt atgtagtttt atgatgatat gtatttaggt tatttttata ataattgggt 7560
tgttaagggt tttatatttg ttttttttatt tattaagagt ttttttttttt ttaattttat 7620
gaatgttaat tttttgttat tatagagtat gttttttttta tttaatttta ttttgtttat 7680
gagtatgttg tttagtatgg tgtttttagt agtgataggt gttttgggtt ttagttttaa 7740
tagtttgaat aatttgaata atttgagtag tttgttgttg aattttgtgg tgttgatgtt 7800
tgtttgtttt tatgtgttgt tgatttttttt gtatgtttat agggatatgt gtaatttgag 7860
tttggttagt ttgagattga aagtaaagta gtattttagt tttggttatg ttagtgtgta 7920
gaatttggtt tttaatttga gtgtttgtta gtatgtagtg gattggtttg tgtgagttgt 7980
atttatagtg ttgggatttt aggattttgt tggatggggt aattttgttt ttgaaagatt 8040
gggaattatg ttagaaggtt gtgggtatta aagaaaggga gagaaagaga agttatatag 8100
agaaaaggaa attattgaat taaagagaga gttttttttga ttttaaaggg atgtttttag 8160
tgtttgatat ttttttattat aagtattttt aatagttgta aggatatata tataaataaa 8220
tgtttgattg gatatgatat tttaatatta ttataagttt gttatttttt aagtttagta 8280
ttgttaatat ttaaatgatt gaaaggatgt atatatattg aaatgttaaa ttaattttat 8340
aaaagtagtt gttagtaata ttataatagt gtttttaaag gttaggtttt aaaataaagt 8400
atgttatata gaagtgatta ggatttttttg tttgtgagta agggagtgta tatattaaat 8460
gttatattgt atgtttttaa tatattatta ttattataaa aaatgtgtga atattagttt 8520
tagaatagtt tttttggtgg atgtaatgat gtttttgaaa ttgttatgta taatttattt 8580
tgtgtataat attttgtata atattattgt tttatttttt agtaaatatg aaataaatgt 8640
gttttatttt atgggagtaa aatatattgt atataaattg gtttggattt ttttttttttt 8700
tttttgttat taatttggtt aggatatttt agttattgtt ttttaaataa attagttttt 8760
tttgtttgtt tagttaaata tataaggtag tagtttttat ttaaatttgg tagaaataaa 8820
tgatagttat ttattagaaa ttaaaaagaa aaaaaaaggt atttttgggg gggaaaaggg 8880
ttataaaatt taattttgtt tttttaattt ttttttggtt taaatttaga ggattttatt 8940
atggttagta aataatatga aaaagaaaaa agaagaaaga aatttagtaa gtttattagt 9000
t 9001

<210> 6
<211> 22
<212> DNA

<213> Artificial Sequence

<220>
<223> PRIMER

<400> 6

gtaggggagg gaagtagatg tt 22

<210> 7
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> PRIMER

<400> 7

ttctaatcct cctttccaca ataa 24

<210> 8
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> PROBE

<400> 8

agtcggagtc gggagagcga 20

<210> 9
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> PROBE

<400> 9

agttggagtt gggagagtga aaggaga 27

<210> 10
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> PRIMER CONTROL

<400> 10

tggtgatgga ggaggtttag taagt 25

<210> 11
<211> 27
<212> DNA
<213> Artificial Sequence

<220>

<223> PRIMER CONTROL

<400> 11

aaccaataaa acctactcct cccttaa 27

<210> 12
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> PROBE CONTROL

<400> 12

accaccaccc aacacacaat aacaaacaca 30

<210> 13
<211> 408
<212> DNA
<213> Homo Sapiens

<400> 13

tcctcaactc tgcaggcctg aaagaaggtc acacacgcac gctcacaccc acactccaca 60
cgcctcgtcc caaacaaccc catgaacatt gtcctttgtt ccgtctcttg ggccactttc 120
cctgtcgctt cctcccagcc cgtcctgatt tgctccccaa aagtacgttt ctgtctcccc 180
gctgccctgg cgctcccccc ttgatttatt agggctgccg ggttggcgca gattgctttt 240
tcttctcttc catcccatcc tcccttctgg tcctcctttc cacagtggga gtccgtgctc 300
ctgctcctcg gttggctcct aagtgccccg ccaggtcccc tctcctttcg ctctcccggc 360
tccggctccc gactcttcgg cccgctggca tctgcttccc tcccctgc 408

<210> 14
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> chemically treated genomic DNA (Homo sapiens)

<400> 14

aacatctact tcccctcccct ac 22

<210> 15
<211> 32
<212> DNA
<213> Artificial Sequence

<220>
<223> chemically treated genomic DNA (Homo sapiens)

<400> 15

gttagtagag attttattaa attttattgt at 32

<210> 16
<211> 14
<212> DNA
<213> Artificial Sequence

<220>

<223> chemically treated genomic DNA (Homo sapiens)

<400> 16

ttcggttgcg cggt 14

<210> 17
<211> 16
<212> DNA
<213> Artificial Sequence

<220>
<223> chemically treated genomic DNA (Homo sapiens)

<400> 17

tttggttgtg tggttg 16

<210> 18
<211> 144
<212> DNA
<213> Homo Sapiens

<400> 18

ttctggtcct cctttccaca gtgggagtcc gtgctcctgc tcctcggttg gctcctaagt 60
gccccgccag gtcccctctc ctttcgctct cccggctccg gctcccgact cttcggcccg 120
ctggcatctg cttccctccc ctgc 144

<210> 19
<211> 162
<212> DNA
<213> Homo Sapiens

<400> 19

tggcatctgc ttccctcccc tgcctcgttt ctcgtcgccc ctgctcgctc cccccggcgc 60
tcgcccgggc gctgtgctcg ctcctggatc gccagccgcg cagcgggctc gccggcgccc 120
gcgcgccact gtgcagtgga gtttggtgga atctctgctg ac 162

<210> 20
<211> 2235
<212> DNA
<213> Homo Sapiens

<400> 20

tgggagtccg tgctcctgct cctcggttgg ctcctaagtg ccccgccagg tcccctctcc 60
tttcgctctc ccggctccgg ctcccgactc ttcggcccgc tggcatctgc ttccctcccc 120
tgcctcgttt ctcgtcgccc ctgctcgctc cccccggcgc tcgcccgggc gctgtgctcg 180
ctcctggatc gccagccgcg cagccgggct cggccggccg cccgcgcgcc actgtgcagt 240
ggagtttggt ggaatctctg ctgacgtcac gtcactcccc acacggagta ggagcagagg 300
gaagagagag ggatgagagg gagggagagg agagagagtg cgagaccgag cgagaaagct 360
ggagaggagc agaaagaaac tgccagtggc ggctagattt cggaggcccc agtgcacccg 420
tggactcctt cggaacttgg caccctcagg agccctgcag tcctctcagg cccggctttc 480
gggcgcttgc cgtgcagccg gaggctcggc tcgctggaaa tcgccccggg aagcagtggg 540
acgcggagac agcagctctc tcccggtagc cgataacggg gaatggagac caactgccgc 600
aaactggtgt cggcgtgtgt gcaattaggc gtgcagccgg cggccgttga atgtctcttc 660
tccaaagact ccgaaatcaa aaaggtcgag ttcacggact ctcctgagag ccgaaaagag 720
gcagccagca gcaagttctt cccgcggcag catcctggcg ccaatgagaa agataaaagc 780
cagcagggga agaatgagga cgtgggcgcc gaggacccgt ctaagaagaa gcggcaaagg 840
cggcagcgga ctcactttac cagccagcag ctccaggagc tggaggccac tttccagagg 900
aaccgctacc cggacatgtc cacacgcgaa gaaatcgctg tgtggaccaa ccttacggaa 960

gcccgagtcc gggtttggtt caagaatcgt cgggccaaat ggagaaagag ggagcgcaac 1020
cagcaggccg agctatgcaa gaatggcttc gggccgcagt tcaatgggct catgcagccc 1080
tacgacgaca tgtacccagg ctattcctac aacaactggg ccgccaaggg ccttacatcc 1140
gcctccctat ccaccaagag cttccccttc ttcaactcta tgaacgtcaa ccccctgtca 1200
tcacagagca tgttttcccc acccaactct atctcgtcca tgagcatgtc gtccagcatg 1260
gtgccctcag cagtgacagg cgtcccgggc tccagtctca acagcctgaa taacttgaac 1320
aacctgagta gcccgtcgct gaattccgcg gtgccgacgc ctgcctgtcc ttacgcgccg 1380
ccgactcctc cgtatgttta tagggacacg tgtaactcga gcctggccag cctgagactg 1440
aaagcaaagc agcactccag cttcggctac gccagcgtgc agaacccggc ctccaacctg 1500
agtgcttgcc agtatgcagt ggaccggccc gtgtgagccg cacccacagc gccgggatcc 1560
taggaccttg ccggatgggg caactccgcc cttgaaagac tgggaattat gctagaaggt 1620
cgtgggcact aaagaaaggg agagaaagag aagctatata gagaaaagga aaccactgaa 1680
tcaaagagag agctcctttg atttcaaagg gatgtcctca gtgtctgaca tctttcacta 1740
caagtatttc taacagttgc aaggacacat acacaaacaa atgtttgact ggatatgaca 1800
ttttaacatt actataagct tgttattttt taagtttagc attgttaaca tttaaatgac 1860
tgaaaggatg tatatatatc gaaatgtcaa attaatttta taaaagcagt tgttagtaat 1920
atcacaacag tgtttttaaa ggttaggctt taaaataaag catgttatac agaagcgatt 1980
aggatttttc gcttgcgagc aagggagtgt atatactaaa tgccacactg tatgtttcta 2040
acatattatt attattataa aaaatgtgtg aatatcagtt ttagaatagt ttctctggtg 2100
gatgcaatga tgtttctgaa actgctatgt acaacctacc ctgtgtataa catttcgtac 2160
aatattattg ttttactttt cagcaaatat gaaacaaatg tgttttattt catgggagta 2220
aaatatactg catac 2235

<210> 21
<211>
<212> protein
<213> polypeptide Sequence

<220>
<223> amino acid sequence (Homo sapiens)

<400> 450

METNCRKLVSACVQLGVQPAAVECLFSKDSEIKKVEFTDSPESRKEAASSKFFP
RQHPGANEKDKSQQGKNEDVGAEDPSKKKRQRRQRTHFTSQQLQELEATFQRNR
YPDMSTREEIAVWTNLTEARVRVWFKNRRAKWRKRERNQQAELCKNGFGPQFNG
LMQPYDDMYPGYSYNNWAAKGLTSASLSTKSFPFFNSMNVNPLSSQSMFSPPNS
ISSMSMSSSMVPSAVTGVPGSSLNSLNNLNNLSSPSLNSAVPTPACPYAPPTPP
YVYRDTCNSSLASLRLKAKQHSSFGYASVQNPASNLSACQYAVDRPV 450

<210> 22
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> chemically treated genomic DNA (Homo sapiens)

<400> 22

gtaggggagg gaagtagatg t 21

<210> 23
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> chemically treated genomic DNA (Homo sapiens)

<400> 23

tcctcaactc tacaaaccta aaa 23

<210> 24
<211> 16
<212> DNA
<213> Artificial Sequence

<220>
<223> chemically treated genomic DNA (Homo sapiens)

<400> 24

agtcgggaga gcgaaa 16

<210> 25
<211> 16
<212> DNA
<213> Artificial Sequence

<220>
<223> chemically treated genomic DNA (Homo sapiens)

<400> 25

agttgggaga gtgaaa 16

<210> 26
<211> 17
<212> DNA
<213> Artificial Sequence

<220>
<223> chemically treated genomic DNA (Homo sapiens)

<400> 26

aagagtcggg agtcgga 17

<210> 27
<211> 17
<212> DNA
<213> Artificial Sequence

<220>
<223> chemically treated genomic DNA (Homo sapiens)

<400> 27

aagagttggg agttgga 17

<210> 28
<211> 16
<212> DNA
<213> Artificial Sequence

<220>
<223> chemically treated genomic DNA (Homo sapiens)

<400> 28

ggtcgaagag tcggga 16

<210> 29
<211> 16

<212> DNA
<213> Artificial Sequence

<220>
<223> chemically treated genomic DNA (Homo sapiens)

<400> 29

ggttgaagag ttggga 16

<210> 30
<211> 16
<212> DNA
<213> Artificial Sequence

<220>
<223> chemically treated genomic DNA (Homo sapiens)

<400> 30

atgttagcgg gtcgaa 16

<210> 31
<211> 16
<212> DNA
<213> Artificial Sequence

<220>
<223> chemically treated genomic DNA (Homo sapiens)

<400> 31

tagtgggttg aagagt 16

<210> 32
<211> 24
<212> DNA
<213> Homo Sapiens

<400> 32

gaaaggcaga gtcataacag gaag 24

<210> 33
<211> 26
<212> DNA
<213> Homo Sapiens

<400> 33

taaggataga gtgatttcca agaaag 26

<210> 34
<211> 24
<212> DNA
<213> Homo Sapiens

<400> 34

cttatctgac aacaagcgag tatg 24

<210> 35
<211> 25

<212> DNA
<213> Homo Sapiens

<400> 35

caattattgt attgtagcat cggag 25